

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 13:13:40 ; Search time 3339 Seconds

(without alignments)
6788.981 Million cell updates/sec

Title: US-09-669-817A-4

Perfect score: 523

Sequence: 1 atcgcacagcagatcacca.....gaccgcaagtggaagtcgga 523

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vt.*

15: em_da.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vt.*

30: em_hcg_hum.*

31: em_hcg_inv.*

32: em_hcg_other.*

33: em_hcg_mus.*

34: em_hcg_pln.*

35: em_hcg_rnd.*

36: em_hcg_mam.*

37: em_hcg_vrt.*

38: em_sy.*

39: em_hsgo_hum.*

40: em_hsgo_mus.*

41: em_hsgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515	98.5	1819	AKI03829	AKI03829 Oryza sat
2	515	98.5	194509	AP001129	AP001129 Oryza sat
3	467.6	89.4	1464	AK060638	AK060638 Oryza sat
4	130.2	24.9	1063	AKI20505	AKI20505 Oryza sat
5	130.2	24.9	1548	AK063929	AK063929 Oryza sat
6	130.2	24.9	1665	AKI02953	AKI02953 Oryza sat
7	130.2	24.9	110989	AP004049	AP004049 Oryza sat
8	130.2	24.9	135189	AP004851	AP004851 Oryza sat
9	109.2	20.9	7218	166494	Sequence 14
10	108.2	20.7	63082	AC022653	AC022653 Homo sapi
11	107	20.5	303091	AC084799	AC084799 Mus muscu
12	106.6	20.4	265537	AC087228	AC087228 Mus muscu
13	106	20.3	69722	ACI00560	ACI00560 Mus muscu
14	106	20.3	252689	AC079433	AC079433 Mus muscu
15	105.8	20.2	3281	ACI43602	ACI43602 Macaca mu
16	105.4	20.2	203664	ACI39759	ACI39759 Mus muscu
17	105.2	20.1	62649	AC022552	AC022552 Homo sapi
18	105.2	20.1	81417	AC023265	AC023265 Homo sapi
19	105	20.1	222895	AC084065	AC084065 Mus muscu
20	104.8	20.0	60885	ACI00962	ACI00962 Mus muscu
21	104.8	20.0	164751	AP003567	AP003567 Danio rer
22	104.8	20.0	176734	AP003682	AP003682 Oryza sat
23	104.6	20.0	65505	AC027799	AC027799 Homo sapi
24	104.6	20.0	300695	AC079431	AC079431 Mus muscu
25	104.4	20.0	61223	ACI30347	ACI30347 Homo sapi
26	104.4	20.0	78220	AC023212	AC023212 Homo sapi
27	104.4	20.0	80291	ACI01154	ACI01154 Mus muscu
28	104.2	19.9	262980	AC098981	AC098981 Rattus no
29	104	19.9	126543	ACI43806	ACI43806 Macaca mu
30	104	19.9	261760	ACI34459	ACI34459 Mus muscu
31	103.8	19.8	239327	AL079409	AL079409 Danio rer
32	103.6	19.8	179206	CNS01056	AL121656 BAC seque
33	103.6	19.8	202083	AC023833	AC023833 Mus muscu
34	103.6	19.8	222707	AC079425	AC079425 Mus muscu
35	103.4	19.8	72847	AC023563	AC023563 Homo sapi
36	103.4	19.8	87283	AC022856	AC022856 Homo sapi
37	103.4	19.8	110737	AC011105	AC011105 Homo sapi
38	103.2	19.7	175146	ACI32466	ACI32466 Mus muscu
39	103	19.7	58483	ACI00352	ACI00352 Mus muscu
40	103	19.7	74138	AC021272	AC021272 Homo sapi
41	102.8	19.7	80133	AC021347	AC021347 Homo sapi
42	102.8	19.7	89994	AC021263	AC021263 Homo sapi
43	102.8	19.7	134940	AC018939	AC018939 Homo sapi
44	102.8	19.7	143733	ACI42786	ACI42786 Macaca mu
45	102.4	19.6	660	CNS0190W	AL112128 Botrytis

ALIGNMENTS

RESULT 1

AKI03829

LOCUS

DEFINITION

AKI03829 Oryza sativa (japonica cultivar-group) cDNA clone:J033148A06, full insert sequence.

ACCESSION

AKI03829

VERSION

AKI03829.1 GI:32989038

KEYWORDS

FLI CDNA; CAP trapper.

SOURCE

ORYZA SATIVA (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaeae; Oryza.

REFERENCE

1 The Rice Full-Length cDNA Consortium, National Institute of

Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Yamamoto, M., and Nakahama, Y.
 FAIS Genome Sequencing & Analysis Group: Oromo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nikura, Y., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arahawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.
 Location/Qualifiers
 1. 1063
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiyar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="J013123A20"

FEATURES

Source

ORIGIN

Query Match 24.9%; Score 130.2; DB 8; Length 1063;
 Best Local Similarity 58.9%; Pred. No. 5.4e-10;
 Matches 269; Conservative 0; Mismatches 173; Indels 15; Gaps 2;
 79 CCACATAGATTTCACCAAGGCTTAAGCCACCTCTCTGAGAGACCCCTTCTCCCTTCC 138
 626 CCACCTTCCCAACCT 567
 139 CACGCGCGCT 198
 566 CTCCT 507
 199 CATCTTCT 258
 506 GCTTCACT 447
 259 CCACG-----CCGACCACTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 306
 446 ACCTTGAACCTTCT 387
 307 TCACCGACCGGCT 366
 386 GCGATCT 327
 367 CCACGAG 426
 326 CCCACGAG 267
 427 CCGAG 486
 266 CCCACGAG 210
 487 CCTACGAG 523
 209 CCAAG 173

RESULT 5
 AK063929
 LOCUS
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:001-123-c12, full

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, T., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Oota, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES
SOURCE
Location/Qualifiers
1..1548
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-123-C12"

Query Match 24.9%; Score 130.2; DB 8; Length 1548;
Best Local Similarity 58.9%; Pred. No. 5e-10;
Matches 269; Conservative 0; Mismatches 173; Indels 15; Gaps 2;

ORIGIN

Query 79 CCACCATGAGTTCACACACGTTAGCCGACCCCTCCGAGACCCCTTCTCCCTCC 138
Db 72 CCACCTTCCCACT 131
Qy 139 CACGCGCGTCT 198
Db 132 CTCCT 191
Qy 199 CATCTTCCGCT 258
Db 192 GCTTCACT 251
Qy 259 CCACG-----CCGACCACTAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 306
Db 252 ACCTTGAACCT 311
Qy 307 TCACCGACCGGCT 366
Db 312 GCGATCCT 371
Qy 367 CCACCGACGAGAGTGAACCTATGTCAACCGATCCGCGCATGTGATGACCA 426
Db 372 CCCGACGAGCAATATACCTCTACGAGCGCGATCGCGGAGGAGATCAAGTGA 431
Qy 427 CCGAGACACGCT 486
Db 432 CCCACGACCAACCC---GACGCGCAGCCGAGCCTTCAAGTGGAGGCCCAATG 488
Qy 487 CCTACGACGATGATTCGACCGCAAGTGAAGTGGGA 523
Db 489 CCAACGACGATGCTTCGACCGCAAGTGAAGTGGGA 525

RESULT 6
AKI02953
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AKI02953 1565 bp mRNA linear PLN 24-JUL-2003
Oryza sativa (japonica cultivar-group) cDNA clone:J03115E12, full
insert sequence.
AKI02953
AKI02953.1 GI:32988162
Full_CDNA; CAP trapper.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE
AUTHORS

Enthartoidae; Oryzae; Oryza.
1 The Rice Full-length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-length cDNA Project team:
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, T., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Nishikawa, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C.,
Chen, X., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Oono, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tanoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikawa, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Mura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Oota, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M. and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)

TITLE
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 1565)

12869764

22752273

TITLE
JOURNAL

Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki
305-8602 Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT

This clone is one of the 28K full-length cDNA clones from japonica
rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team_Kikuchi_S_Satoh_K_Nagata_T_Kawagashira_N_Doi_K_Kishimoto_N_Yazaki_T_Ishikawa_M_Yamada_H_Ooka_H_Hotta_I_Kojima_K_Nishikawa_T_Ohneda_E_Yanagi_W_Suzuki_K_Li_C_Chen_X_Shishiki_T_Yoshimura_A_Matsubara_K_and_Murakami_K

FAIS Genome Sequencing & Analysis Group: Oono, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Mura, J.,
Mizuno, K., Narikawa, R., Nishikawa, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Tanoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Nemuyama, T., Miyazaki, A., Murate, M., Nakamura, M., Nishik, K., Nomura, K., Numesaki, R., Ohno, M., Ootao, N., Ota, Y., Satoh, H., Sawai, C., Sakai, K., Sakazume, N., Sato, H., Sasai, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogaue, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takano-Kahira, S., Tanaka, T., Tomaru, A., Toyota, T., Waki, K., Yasunishi, A. and Hayashiaki, Y.

Location/Qualifiers

```

FEATURES
  source
    Location/Qualifiers
      1..1565
        /organism="Oryza sativa (japonica cultivar-group)"
        /mol_type="mRNA"
        /cuiEiAr="Nipponbare"
        /db_xref="taxon:39947"
        /clone="J033115E12"

```

Query Match 24.9%; Score 130.2; DB 8; Length 1565;
Best Local Similarity 58.9%; Pred. No. 4.9e-10;
Matches 269; Conservative 0; Mismatches 173; Indels 15; Gaps 2.

[illegible]

RESULT 7	AP004049	LOCUS	DEFINITION
	AP004049	110989 bp	DNA
			1 linear
			PLN 27-NOV-2003
			Oriza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
			BAC clone:OU1212_C06, complete sequence.

TITLE Direct Submission
JOURNAL Submitted (15-AUG-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-7, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>
Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Nov 26, 2003 this sequence version replaced gi:15208417.
The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. The nucleotide sequence of this BAC clone was generated by
combining Monsanto and R&P-Japan sequencing data.

```

FEATURES
Source
Location/Qualifiers
1..110989
/organism="Oryza sativa (japonica cultivar-group)"
/mol type="genomic DNA"
/cultivar="Nippondare"
/db xref="taxon:39947"
/chromosome="2"
/clone="OJ1212_C06"

```

Query Match	24.9%;	Score 130.2;	DB 8;	Length 110983;
Best Local Similarity	58.9%;	Pred. No. 1.9e-10;		
Matches 269; Conservative	0;	Mismatches 173;	Indels 15;	Gaps

[illegible]

RESULT	7
LOCUS	AP004049
DEFINITION	AP004049 110989 bp DNA linear PLN 27-NOV-2003
ACCESSION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
VERSION	BAC clone:OU1212_C06, complete sequence.
KEYWORDS	AP004049.2 GI:38564198
SOURCE	Htg.
ORGANISM	Oryza sativa (japonica cultivar-group)
	Oryza sativa (japonica cultivar-group)
	Euryotia; Vitridiplanetae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
	Eurharoidae; Oryzaceae; Oryza.
REFERENCE	1 Sasaki, T., Matsumoto, T. and Yamamoto, K.
AUTHORS	Oryza sativa niponbare (GAS) genomic DNA, chromosome 2, BAC
TITLE	clone:OU1212_C06
JOURNAL	Published Only In Database (2001)
REFERENCE	2 (bases 1 to 110989)
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.

TITLE
JOURNAL
COMMENT

Landers, T., Leheczy, J., Levine, R., Lien, C., Liu, G., Locke, K.,
Macdonald, P., Margis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910643.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4854
Center clone name: 29_O_5

* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 734: contig of 734 bp in length
* 735 834: gap of 100 bp in length
* 835 1536: contig of 722 bp in length
* 1537 1556: gap of 100 bp in length
* 1557 2352: contig of 696 bp in length
* 2353 2452: gap of 100 bp in length
* 2453 3148: contig of 696 bp in length
* 3149 3248: gap of 100 bp in length
* 3249 3945: contig of 687 bp in length
* 3946 4045: gap of 100 bp in length
* 4046 4728: contig of 683 bp in length
* 4729 4829: gap of 100 bp in length
* 4829 5548: contig of 720 bp in length
* 5549 5648: gap of 100 bp in length
* 5649 6362: contig of 714 bp in length
* 6363 6462: gap of 100 bp in length
* 6463 7171: contig of 709 bp in length
* 7172 7271: gap of 100 bp in length
* 7272 7997: contig of 726 bp in length
* 7998 8097: gap of 100 bp in length
* 8098 8835: contig of 738 bp in length
* 8836 8935: gap of 100 bp in length
* 8935 9664: contig of 729 bp in length
* 9665 9764: gap of 100 bp in length
* 9765 10515: contig of 751 bp in length
* 10516 10615: gap of 100 bp in length
* 10616 11429: contig of 814 bp in length
* 11430 11529: gap of 100 bp in length
* 11530 12294: contig of 765 bp in length
* 12295 12394: gap of 100 bp in length
* 12395 13176: contig of 782 bp in length
* 13177 13276: gap of 100 bp in length
* 13277 13891: contig of 715 bp in length
* 13892 14091: gap of 100 bp in length
* 14092 14797: contig of 706 bp in length
* 14798 14897: gap of 100 bp in length
* 14898 15607: contig of 710 bp in length

15608 15707: gap of 100 bp in length
* 15708 16432: contig of 725 bp in length
* 16433 16532: gap of 100 bp in length
* 16533 17268: contig of 736 bp in length
* 17269 17368: gap of 100 bp in length
* 17369 18105: contig of 737 bp in length
* 18106 18205: gap of 100 bp in length
* 18206 18888: contig of 683 bp in length
* 18889 18988: gap of 100 bp in length
* 18989 19702: contig of 714 bp in length
* 19703 19802: gap of 100 bp in length
* 19803 20516: contig of 714 bp in length
* 20517 20616: gap of 100 bp in length
* 20617 21330: contig of 714 bp in length
* 21331 21430: gap of 100 bp in length
* 21431 22143: contig of 713 bp in length
* 22144 22243: gap of 100 bp in length
* 22244 22950: contig of 707 bp in length
* 22951 23050: gap of 100 bp in length
* 23051 23785: contig of 735 bp in length
* 23786 23885: gap of 100 bp in length
* 23886 24606: contig of 721 bp in length
* 24607 24706: gap of 100 bp in length
* 24707 25447: contig of 741 bp in length
* 25448 25547: gap of 100 bp in length
* 25548 26276: contig of 729 bp in length
* 26277 26376: gap of 100 bp in length
* 26377 27107: contig of 731 bp in length
* 27108 27207: gap of 100 bp in length
* 27208 27926: contig of 719 bp in length
* 27927 28026: gap of 100 bp in length
* 28027 28737: contig of 711 bp in length
* 28738 28837: gap of 100 bp in length
* 28838 29559: contig of 722 bp in length
* 29560 29659: gap of 100 bp in length
* 29660 30374: contig of 715 bp in length
* 30375 30474: gap of 100 bp in length
* 30475 31184: contig of 710 bp in length
* 31185 31284: gap of 100 bp in length
* 31285 32011: contig of 727 bp in length
* 32012 32111: gap of 100 bp in length
* 32112 32836: contig of 725 bp in length
* 32837 32936: gap of 100 bp in length
* 32937 33621: contig of 685 bp in length
* 33622 33721: gap of 100 bp in length
* 33722 34444: contig of 723 bp in length
* 34445 34544: gap of 100 bp in length
* 34545 35253: contig of 709 bp in length
* 35254 35353: gap of 100 bp in length
* 35354 36069: contig of 716 bp in length
* 36070 36169: gap of 100 bp in length
* 36170 36878: contig of 709 bp in length
* 36879 36978: gap of 100 bp in length
* 36979 37677: contig of 699 bp in length
* 37678 37777: gap of 100 bp in length
* 37778 38489: contig of 712 bp in length
* 38490 38589: gap of 100 bp in length
* 38590 39297: contig of 708 bp in length
* 39298 39397: gap of 100 bp in length
* 39399 40125: contig of 728 bp in length
* 40126 40225: gap of 100 bp in length
* 40226 40945: contig of 720 bp in length
* 40946 41045: gap of 100 bp in length
* 41046 41767: contig of 722 bp in length
* 41768 41867: gap of 100 bp in length
* 41868 42603: contig of 736 bp in length
* 42604 42703: gap of 100 bp in length
* 42704 43420: contig of 717 bp in length
* 43421 43520: gap of 100 bp in length
* 43521 44223: contig of 703 bp in length
* 44224 44323: gap of 100 bp in length
* 44324 45041: contig of 718 bp in length
* 45042 45141: gap of 100 bp in length

```

* 45142 45855: contig of 714 bp in length
* 45956 45955: gap of 100 bp
* 45956 46679: contig of 724 bp in length
* 46680 46779: gap of 100 bp
* 46780 47482: contig of 703 bp in length
* 47483 47582: gap of 100 bp
* 47583 48317: contig of 735 bp in length
* 48318 48417: gap of 100 bp
* 48418 49147: contig of 730 bp in length
* 49148 49247: gap of 100 bp
* 49248 49967: contig of 720 bp in length
* 49968 50067: gap of 100 bp
* 50068 50772: contig of 705 bp in length
* 50773 50872: gap of 100 bp
* 50873 51568: contig of 696 bp in length
* 51569 51668: gap of 100 bp
* 51669 52399: contig of 731 bp in length
* 52400 52499: gap of 100 bp
* 52500 53224: contig of 725 bp in length
* 53225 53324: gap of 100 bp
* 53325 54013: contig of 689 bp in length
* 54014 54113: gap of 100 bp
* 54114 54844: contig of 731 bp in length
* 54845 54944: gap of 100 bp
* 54945 55660: contig of 716 bp in length
* 55661 55760: gap of 100 bp
* 55761 56490: contig of 730 bp in length
* 56491 56590: gap of 100 bp
* 56591 57333: contig of 743 bp in length
* 57334 57433: gap of 100 bp
* 57434 58161: contig of 728 bp in length
* 58162 58261: gap of 100 bp

Query Match
Best Local Similarity 47.6%; Score 108.2; DB 2; Length 63082;
Matches 221; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

16 CACCAATCTCTCTTCACTCCCACTCCAGCGAGTGAAGAGTCCACCGCCG 75
36038 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35979
76 CGGCGACCATGAGTTCTACACCGCTTACCGCACTCTCTGACGACCCCTTCTCCCT 135
35978 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35919
136 TCCGACCGCGCTCTCTCTCTCTGCGCCCTTCTCTTCCCAACCGCGGTTCTCTCACT 195
35918 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35859
196 GCCCATCTTGCGCCCTGCACTCCCGCTTGCTGCTGCACTTCCATCTCCACCCCTTC 255
35858 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35799
256 TCCGCAAGCGGACCATGAGTCTCTCTCTGACCCCTTCTCTCTCCACACCTTACCGACC 315
35798 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35739
316 GCGCTCTCCAACTGAGTGTGCGCGCGCGCGCGCCACACCCAGGCGCCACAGCA 375
35738 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35679
376 GGAAGTGCACCTATGTACCGAGTCCACCGGCGCATGTCAAGTGAACACCGAGACA 435
35678 CCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 35619
436 AGCTTCGCGCGCGGAGAGGTGTCAAGTGGAGGCTGAGCTC 479
35618 ANCTTCCACCTGTGTGCGCGCCCTCCAGGCTCCTGTGTCTC 35575

```

RESULT 11
AC084799/c 303091 bp DNA linear HTG 17-NOV-2000

```

DEFINITION
Mus musculus chromosome 16 clone RP23-197M9, WORKING DRAFT
SEQUENCE, 101 unordered pieces.
ACCESSION
AC084799.1 GI:11192127
VERSION
HTG; HTGS; PHASE1; HTGS; DRAFT.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 303091)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 303091)
DOE Joint Genome Institute.
Direct Submission
Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-23_197M9
-----
Summary Statistics
Consensus quality: 152568 bases at least Q40
Consensus quality: 175579 bases at least Q30
Consensus quality: 186349 bases at least Q20
Estimated insert size: 198300; agarose-1p estimation
Estimated insert size: 293091; sum-of-contigs estimation
Quality coverage: 4.85 in Q20 bases; agarose-1p estimation
Quality coverage: 3.28 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 101 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1116: contig of 1116 bp in length
1117 1216: gap of unknown length
1217 2536: contig of 1520 bp in length
2537 2636: gap of unknown length
2637 4057: contig of 1421 bp in length
4058 4158: gap of unknown length
4159 5245: contig of 1088 bp in length
5246 5345: gap of unknown length
5346 6461: contig of 1116 bp in length
6462 6561: gap of unknown length
6562 7725: contig of 1164 bp in length
7726 7826: gap of unknown length
7827 9143: contig of 1318 bp in length
9144 9243: gap of unknown length
9244 10492: contig of 1249 bp in length
10493 10592: gap of unknown length
10593 11666: contig of 1074 bp in length
11667 11766: gap of unknown length
11767 12950: contig of 1184 bp in length
12951 13050: gap of unknown length
13051 14285: contig of 1235 bp in length
14286 14385: gap of unknown length
14386 15478: contig of 1093 bp in length
15479 15578: gap of unknown length
15579 16779: contig of 1201 bp in length
16780 16779: gap of unknown length
16780 18278: contig of 1399 bp in length
18279 18379: gap of unknown length
18379 19452: contig of 1074 bp in length
19453 19552: gap of unknown length

```

19553	21012:	contig of 160 bp in length
21013	21112:	gap of unknown length
21113	22169:	contig of 1051 bp in length
22170	22267:	gap of unknown length
23368	22367:	contig of 1098 bp in length
23468	22467:	gap of unknown length
23469	22559:	contig of 1092 bp in length
24560	24659:	gap of unknown length
24660	25825:	contig of 1166 bp in length
25826	25925:	gap of unknown length
25926	27036:	contig of 1111 bp in length
27037	27136:	gap of unknown length
27137	28245:	contig of 1109 bp in length
28246	28346:	gap of unknown length
28346	29556:	contig of 1311 bp in length
29557	29757:	gap of unknown length
30884	30883:	contig of 1127 bp in length
30884	32125:	gap of unknown length
32126	33225:	gap of unknown length
33226	33497:	contig of 1172 bp in length
33498	33497:	gap of unknown length
33498	34645:	contig of 1148 bp in length
34646	34745:	gap of unknown length
34746	35694:	contig of 1149 bp in length
35695	35994:	gap of unknown length
35995	37156:	contig of 1162 bp in length
37157	37256:	gap of unknown length
37257	38396:	contig of 1140 bp in length
38397	38496:	gap of unknown length
38497	39570:	contig of 1074 bp in length
39571	39670:	gap of unknown length
39671	40852:	contig of 1182 bp in length
40853	40952:	gap of unknown length
40953	42116:	contig of 1164 bp in length
42117	42216:	gap of unknown length
42217	43372:	contig of 1156 bp in length
43373	44372:	gap of unknown length
44373	44603:	contig of 1131 bp in length
44604	44703:	gap of unknown length
44704	45924:	contig of 1221 bp in length
45925	46024:	gap of unknown length
46025	47174:	contig of 1150 bp in length
47175	47774:	gap of unknown length
47775	48744:	contig of 1470 bp in length
48745	48844:	gap of unknown length
48845	49969:	contig of 1125 bp in length
49970	50069:	gap of unknown length
50070	51135:	contig of 1066 bp in length
51136	51235:	gap of unknown length
51236	52690:	contig of 1455 bp in length
52691	53200:	gap of unknown length
53201	54020:	gap of unknown length
54021	55101:	contig of 1081 bp in length
55102	55201:	gap of unknown length
55202	56324:	contig of 1123 bp in length
56325	56424:	gap of unknown length
56425	57583:	contig of 1159 bp in length
57584	57683:	gap of unknown length
57684	58622:	contig of 1139 bp in length
58623	58922:	gap of unknown length
58923	60547:	contig of 1625 bp in length
60548	60647:	gap of unknown length
60648	61743:	contig of 1096 bp in length
61744	61843:	gap of unknown length
61844	62998:	contig of 1155 bp in length
62999	63098:	gap of unknown length
63099	64220:	contig of 1122 bp in length
64221	64320:	gap of unknown length
64321	65398:	contig of 1078 bp in length
65399	65498:	gap of unknown length
65499	66762:	contig of 1264 bp in length

	*	66763	:	gap of unknown length	
	*	66823	:	contig of 163 bp in length	
	*	68026	:	gap of unknown length	
	*	68126	:	contig of 1275 bp in length	
	*	69401	:	gap of unknown length	
	*	69501	:	contig of 1218 bp in length	
	*	70719	:	gap of unknown length	
	*	70819	:	contig of 1177 bp in length	
	*	71996	:	gap of unknown length	
	*	72096	:	contig of 1137 bp in length	
	*	73233	:	gap of unknown length	
	*	73333	:	contig of 1144 bp in length	
	*	74477	:	gap of unknown length	
	*	74577	:	contig of 1287 bp in length	
	*	75864	:	gap of unknown length	
	*	75964	:	contig of 1058 bp in length	
	*	77022	:	gap of unknown length	
	*	77122	:	contig of 1091 bp in length	
	*	78213	:	gap of unknown length	
	*	78313	:	contig of 1112 bp in length	
	*	79425	:	gap of unknown length	
	*	79525	:	contig of 1093 bp in length	
	*	80618	:	gap of unknown length	
	*	80717	:	gap of unknown length	
	*	81078	:	contig of 1087 bp in length	
	*	81804	:	gap of unknown length	
	*	81905	:	contig of 1103 bp in length	
	*	83008	:	gap of unknown length	
	*	83108	:	contig of 1060 bp in length	
	*	84168	:	gap of unknown length	
	*	84268	:	contig of 1080 bp in length	
	*	85348	:	gap of unknown length	
	*	85448	:	contig of 1074 bp in length	
	*	86522	:	gap of unknown length	
	*	86622	:	contig of 1123 bp in length	
	*	87745	:	gap of unknown length	
	*	87845	:	contig of 1121 bp in length	
	*	88965	:	gap of unknown length	
	*	88966	:	contig of 1111 bp in length	
	*	90177	:	gap of unknown length	
	*	90276	:	contig of 1003 bp in length	
	*	91279	:	gap of unknown length	
	*	91280	:	contig of 1184 bp in length	
	*	91380	:	contig of 12563 bp in length	
	*	92564	:	gap of unknown length	
	*	92664	:	contig of 1025 bp in length	
	*	93689	:	gap of unknown length	
	*	93789	:	contig of 1146 bp in length	
	*	94935	:	gap of unknown length	
	*	95035	:	contig of 1077 bp in length	
	*	96112	:	gap of unknown length	
	*	96212	:	contig of 2141 bp in length	
	**	98453	:	gap of unknown length	
	*	98453	:	contig of 1193 bp in length	
	*	99646	:	gap of unknown length	

Query Match 20.5%; Score 107; DB 2; Length 303091;
Best Local Similarity 51.9%; P-adj. No. 3.7e-07;
Matches 182; Conservative 0; Mismatched 169; Indels 0; Gaps 0;

Df	44304	CCGCCTTACACTCCCACTCGAGGTGAGTGAGCAGAGCTTCACACGGCGCGGAGCACCC	83
Df	44304	CCCCNNCCCCCCCCCCCCCNCCCNCNCCCCCNCCCCCNCCCCCNCCCCCNCCCCCNCCC	44245
Oy	84	ATGAGTTCTACACCAGCGTAAAGCCCACTTCCTGACGACGCCCTTTCCCTCCACCG	143
Df	44244	CCCNCCCCCCCCCCCCCCCCCNCCCNCNCCCCCNCCCNCNCCCCCCCCCCCCCNCCCC	44185
Oy	144	CCGNCCTCTCTCTCGAGCCCTTCTCTCTCCACACGGCGGTTTCTCTCACTGCACATT	203
Df	44184	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCNCCNCCCCCCCCCCCCCNCCN	44125
Oy	204	TTCGCTCTGAGCTCCGCTTCTGCTGTGAACCTTTTCATCTGACGCGCTTCTCCGACG	263
Df	44124	CCCCCCCCCCCCCCCCCCCCCNCCCNCNCCCCCCCCCNCCNCCNCCCCCCCCCNCCC	44065

FEATURES	
*	Location/Qualifiers
*	
*	127566 140569: contig of 12774 bp in length
*	140370 140469: gap of unknown length
*	140470 151583: contig of 1113 bp in length
*	151583 151682: gap of unknown length
*	151683 164408: contig of 17726 bp in length
*	164409 164508: gap of unknown length
*	164509 177104: contig of 1536 bp in length
*	177105 177204: gap of unknown length
*	177205 189144: contig of 11940 bp in length
*	189145 189244: gap of unknown length
*	189245 206994: contig of 17750 bp in length
*	206995 227094: gap of unknown length
*	227095 227854: contig of 20660 bp in length
*	227855 227854: gap of unknown length
*	227855 265537: contig of 37683 bp in length

ORIGIN

Query Match	20.4%;	Score 106.6;	DB 2;	Length 265537;
Best Local Similarity	49.2%;	Pred. No. 4.4e-07;		
Matches 190;	Conservative 0;	Mismatches 196;	Indels 0;	Gaps 0;

[illegible]

RESULT 13	
AC100560	
LOCUS	AC100560 69722 bp DNA linear HTG 22-NOV-2001
DEFINITION	Mus musculus clone RP23-154B17, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC100560
VERSION	AC100560.1 GI:17047926
KEYWORDS	HTG; HTGS PHASE0.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 69722)
REFERENCE	Birren, B., Linton, L., Nusbaw, C. and Lander, E.
AUTHORS	
TITLE	Mus musculus, clone RP23-154B17
JOURNAL	unpublished

REFERENCE
AUTHORS

REFERENCE
AUTHORS
2 (bases 1 to 69722))
Birtan, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguski, V., Boulanger, B.,
Brown, A., Camarata, J., Campiano, A., Chang, J., Charao, B.,
Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Deariellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farrow, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Gargan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hume, M., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacombe, K.,
Lamaatere, R., Landers, T., Lemocky, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margus, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneilly, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nordh, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, U., Peterson, K., Phunkhang, P., Piere, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Romeh, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travs, N., Triggilo, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:

Db 4997 CCCCCC 4992

RESULT 3

ABK98592/c
ID ABK98592 standard; DNA; 12739 BP

AC ABK98592;

DT	07-AUG-2003	(revised)
DT	21-OCT-2002	(first entry)

DE Vector pPEPF1 containing Xy1R/Xy1O/CP25 sequences.

KW ds; promoter; gram positive bacteria; fusion promoter; T5; CP25; p32;
 KW p59; p1p2; Pl; xylo; teco; trpo; marO; lambdaclacO; cellular proliferation;
 KW antibiotic; vector.

OS Eubacteria.

OS Bacteriophage lambda.
OS Escherichia coli.

OS Synthetic.

OS Synthetic.

PN WO200251982-A2

PD 04-JUL-2002

PF 21-DEC-2001; 2001WO-US050250

PR 27-DEC-2000; 2000US-0259434P

PR 06-SEP-2001; 2001US-00948993.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Wall D, Gross M;

DR WPI; 2002-575374/61.
XX

PT Isolated nucleic acid comprises bacterial promoters modified to have altered activity in at least one gram-positive organism, e.g. *Bacillus anthracis* or *Clostridium botulinum*, useful for regulating gene expression in bacteria..

PS Example 1; Page 206-209; 246pp; English

CC The invention relates to an isolated nucleic acid comprising a fusion
CC promoter comprising at least one promoter that is modified to have
CC altered activity in at least one Gram-positive organism, or comprising
CC 35, CP25, P32, P59, P42 or PL linked to at least one operator consisting
CC of xy10, tccO, trpO, malO or lamBac10, where at least one operator is
CC positioned so binding of a repressor to an operator represses
CC transcription from the fusion promoter. Also included are vectors and
CC host cells comprising the fusion promoters, a method of identifying genes
CC involved in cellular proliferation or required for proliferation of a
CC prokaryotic cell using the vector, a method of identifying compounds that
CC inhibit the proliferation of a prokaryotic cell using the vector, a
CC method of identifying a compound that reduces the activity or level of a
CC gene product required for proliferation of a cell using the vector, a
CC compound identified by the methods, a method of inhibiting the activity
CC or expression of a gene in an operon required for proliferation using the
CC vector, manufacturing an antibiotic comprising using the vector or cell
CC and identifying a nucleic acid with promoter activity in *Enterococcus*
CC faecalis. The fusion promoters are useful for regulating nucleic acid or
CC polypeptide expression, particularly for regulating gene expression in
CC bacteria and for identifying proliferation-regulated genes or molecules
CC with potential antibiotic activity. The modified promoters are also
CC useful for replacing endogenous promoters to create cells with specific
CC regulatable genes. The present sequence is vector (or fragment)
CC incorporating a fusion promoter sequence of the invention. (Updated on 07
CC AUG-2003 to correct OS field.)

Sequence 12739 BP; 3543 A; 2572 C; 3211 G; 3413 T; 0 U; 0 Other;

Query Match	19.5%	Score 102;	DB 6;	length 12739;
Best Local Similarity	54.9%	Pred. No. 3.8e-10;		
Matches 201; Conservative	0;	Mismatches 165;	Indels 0;	Gaps 0;

QY	6	ATCAGAGATACCAAGTCTCTTCAACTGCCACTCAAGGAGTACGAGAGCT	65
Db	5363	ATGGTGTGAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	5304
QY	66	CCACGGCGCGGGCCACCATGATTCTCACACCGTTACGCGCCACTCTCGACGACCC	125
Db	5303	CC	5244
QY	126	TTCTTCCCTTCCACGCGCGTCTCTCTCTGCCCCCTTCTTCCCAACCGCGCTT	185
Db	5243	CC	5184
QY	186	TCCTCAACGTGGCAATTTTGAGCTGAGTCCGCCCTTGCGTGTACGCTTTCATATC	245
Db	5183	CC	5124
QY	246	CACCCCTTCTCCACGCGCAGCACTAGTCTCTCTGACCCCTTCTCTCAACAC	305
Db	5123	CC	5064
QY	306	CTCACGACGCGTCTCCAACTGAGCTGCGCGCGCGCGCCCAACCCCAAG	365
Db	5063	CC	5004
QY	366	CCCAACC 371	
Db	5003	CCCCCC 4998	

RESULT 4
ACD13843

ID ACD13843 standard; DNA; 12739 BP

AC ACD13843;

DT 15-AUG-2003 (First entry)

Plasmid pPEPF1 containing a *Xyl15-CP25* fusion promoter

KM Promoter; ds; gram positive bacteria; Staphylococcus aureus; plasmid

cellular proliferation

05 Lactococcus lactis
05 Synthetic

AA
PN
HS2003027286-A1

06-FEB-2003,

21-DEC-2001; 2001US-00032393

06-SEP-2000; 2000US-0230335P

XX

PA (WALL/) WALL D.

XX

Figure 1

PT New isolated nucleic acid comprising a fusion promoter having at least one promoter that is modified to have altered activity in at least one gram-positive organism, useful for regulating gene expression in bacteria.

PS Example 1; Page 63-68; 142pp; English

D_b 137 CC 78
QY 289 CCTTCTCCTCCACACCCTCACCGAGCGGTCTCCAACTGAGTGGCCCTGGCGGCC 345
D_b 77 CC 18
QY 349 GCGGCCCAACGCC 362
D_b 17 CCCCCCCCCCCCCCCC 4

```

RESULT 8
AAA38185/c
ID      AAA38185 standard; DNA; 320 BP.
vv

```

DT	15-SEP-2003	(revised)
DT	01-SEP-2000	(first entry)
XX		

Primer used in the analysis of a BVDV genome fragment.

KM Primer; bovine viral diarrhoea virus; BVDV; nucleic acid analysis;
KW diagnosis; pathological organism; detect; ss.

OS Pestivirus type 1.

PN WO200020628-A1

PD 13-APR-2000

PF 01-OCT-1999; 99WO-CA000915.

PR 01-OCT-1998; 98US-00165264.

PA (BIOI-) BIO-ID DIAGNOSTIC INC.

PI Vinayagamorthy T.

DR WPI; 2000-303800/26.

PT Nucleic acid analysis methods for simultaneously analyzing multiple nucleic acid regions for diagnosis and differentiation of pathological organisms comprises sequencing the nucleic acids in the reaction mixture

PS Example 2; Page 23; 36pp; English

CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhoea virus (BVD) genome. The primer is used to
CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for
CC simultaneously analysing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the size of the amplified
CC region to be increased. The technique is especially useful when the
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subsets of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC predisposition to a disease. (Updated on 15-SEP-2003 to standardise OS
CC field)

Sequence 320 BP; 4 A; 6 C; 308 G; 2 T; 0 U; 0 Other,

Query Match	Score	DB	Length
18.6%	97.4	3	320

Matches 176; Conservative 0; Mismatches 131; Indels 0; Gaps 0.

65 TCCACCGCGCGGCCACCATGATTTCACACCGGTACGCCACCTCTCTCGACGACCC 120

[illegible]

RESULT 9	
AAZ77506	
ID	AAZ77506 standard; cDNA; 2188 BP

AC AA277506;
xy

DT 10-APR-2000 (first entry
xx

DE Human ovarian tumor cDNA library derived EST fragment 57

XM Expressed sequence tag; EST; human; ovarian tumor; anticancer
 XM gene therapy; treatment; ss.

OS Homo sapiens

PN DE19817557-A1

PD 21-OCT-1999

PF 09-APR-1998; 98DE-01017557

PR 09-APR-1998; 98DE-01017557

PA (META-) METAGEN GES GENOMFORSCHUNG MBH

PI. Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, ...

DR WPI; 1999-591920/51

XXIII

tissues, and derived polypeptides, for

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

33

CC have anticancer activity and are highly expressed in ovarian tumor tissues

CC invention can be used for gene therapy. (A) are used (1) for recombinant

are used (1) to identify agents suitable for treatment of ovarian cancer;

CC gene therapy vectors) and (iii) for generation of specific antibodies.

particular tissue type before comparison of expression patterns. This

CC should reduce the number of failures associated with the fact that ESTS

CC unknown gene, distorting the estimated frequency of occurrence in a

CC Particular tissue: AA277450-Z77572 represent the human ovarian tumor CDNA
CC library derived EST fragments described in the method of the invention
CC and encode the protein fragments represented in AAY76505-Y76638
XX
SQ Sequence 2188 BP; 590 A; 567 C; 434 G; 597 T; 0 U; 0 Other;
Query Match 18.6%; Score 97.2; DB 2; Length 2188;
Best Local Similarity 57.1%; Pred. No. 3e-09;
Matches 177; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
QY 66 CCACCCGCGCGGCGCACATGATCTCTACACACCGTTAGCGCCACCTGCTGACGAGACCC 125
Db 5 CCC 64
QY 126 TTCTTCCCTCTCCACCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
Db 65 CCC 124
QY 186 TCCCTCAGCGGCGCATTTCTTGCGGCTGAGTCCGCTTGCGTGTGAGCCCTTTCATCTC 245
Db 125 CCC 184
QY 246 CACCCCTTCTCTCCACGCGGCGCACATGATCTCTCTCTCTCTCTCTCTCTCTCTCT 305
Db 185 CCC 244
QY 306 CTCACGCGCGGCGTCTCCGACATCTGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGG 365
Db 245 CCC 304
QY 366 CCCACGACGA 375
Db 305 CCCTCCCCCA 314
RESULT 10
ID AAA38184/c
ID AAA38184 standard; DNA; 318 BP.
AC AAA38184;
XX
DT 15-SEP-2003 (revised)
DT 01-SEP-2000 (first entry)
XX
DE Primer used in the analysis of a BVDV genome fragment.
XX
XX Primer; bovine viral diarrhoea virus; BVDV; nucleic acid analysis;
XX diagnosis; pathological organism; detect; ss.
XX
OS Pestivirus type 1.
XX
XX WO200020628-A1.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-CA000915.
XX
XX 01-OCT-1998; 98US-00165264.
XX
XX (BIOI-) BIO-ID DIAGNOSTIC INC.
XX
XX Vinayagamoorthy T;
XX
XX WPI; 2000-303800/26.
XX
XX
XX Nucleic acid analysis methods for simultaneously analyzing multiple
XX nucleic acid regions for diagnosis and differentiation of pathological
XX organisms comprises sequencing the nucleic acids in the reaction mixture.
XX
XX Example 2; Page 23; 36pp; English.
XX
XX This sequence represents a primer used in the analysis of a fragment of
XX the bovine viral diarrhoea virus (BVDV) genome. The primer is used to

CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for
CC simultaneously analyzing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the size of the amplified
CC region to be increased. The technique is especially useful when the usual
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subsets of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC field)
CC predilection to a disease. (Updated on 15-SEP-2003 to standardise OS
CC field)
XX
SQ Sequence 318 BP; 3 A; 3 C; 306 G; 6 T; 0 U; 0 Other;
Query Match 18.2%; Score 95.4; DB 3; Length 318;
Best Local Similarity 57.0%; Pred. No. 6.4e-09;
Matches 174; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 64 CTCACGCGCGGCGGCGCACATGATCTCTACACCGTTAGCGCCACCTCTGACGAGC 123
Db 306 CACCAACCC 247
QY 124 CTTTCTTCCCTTCCACGCGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 183
Db 246 CCC 187
QY 184 TTCTCTCAGCTGCGCATTTCTTGCGGCTGAGTCCGCTTGCTGCTGACCTTTTCATC 243
Db 186 CCC 127
QY 244 TCCACCCCTTCTCTCCACGCGGCGCACATGATCTCTCTCTCTCTCTCTCTCTCTCTCT 303
Db 126 CCC 67
QY 304 CCTCAGCGCGGCTCTCCGACATCTGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCA 363
Db 66 CCC 7
QY 364 GGCCC 368
Db 6 CCCCC 2
RESULT 11
ID AAA38183/c
ID AAA38183 standard; DNA; 320 BP.
AC AAA38183;
XX
XX
DT 15-SEP-2003 (revised)
DT 01-SEP-2000 (first entry)
XX
XX
DE Primer used in the analysis of a BVDV genome fragment.
XX
XX
XX Primer; bovine viral diarrhoea virus; BVDV; nucleic acid analysis;
XX diagnosis; pathological organism; detect; ss.
XX
XX Pestivirus type 1.
XX
XX OS
XX WO200020628-A1.
XX
XX 13-APR-2000.
XX
XX
XX 01-OCT-1999; 99WO-CA000915.
XX
XX 01-OCT-1998; 98US-00165264.
XX
XX (BIOI-) BIO-ID DIAGNOSTIC INC.
XX
XX Vinayagamoorthy T;
XX
XX

XX
DR WPI; 2000-303800/26
XX

PT Nucleic acid analysis methods for simultaneously analyzing multiple
PT nucleic acid regions for diagnosis and differentiation of pathological
PT organisms comprising sequencing the nucleic acids in the reaction mixture

PS Example 2; Page 23; 36pp; English.

CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhoea virus (BVDV) genome. The primer is used to
CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for
CC simultaneously analysing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the site of the amplified
CC region to be increased. The technique is especially useful when the usual
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subtypes of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC predisposition to a disease. (Updated on 15-SEP-2003 to standardise OS
CC field)

SQ Sequence 320 BP; 9 A; 2 C; 305 G; 4 T; 0 U; 0 Other.

Query Match	18.2%	Score 95.2	DB 3	Length 320
Best Local Similarity	57.3%	Pred: No. 7e-09		
Matches 172; Conservative	0	Mismatches 128	Indels 0	Gaps 0

QY 6 GCTCAGCGGCGGGCAACATGAATTTCACACACGTTAGGCCACTCTCGAGAC 122
Db 302 GCCC 243
QY 123 CCCTTCTCCCTTTGCCACGCGGTCTCTCTCTCGCGCTTCTCTGCCACGCGC 182
Db 242 CC 183
QY 183 GTTTCGCACACGAGCCCATCTTGCGCGTCGACATCCCCCTTGCGTGCACCTTTGCAT 243
Db 182 CC 123
QY 243 CTCCACCCTTCTCCGCCACGCGACAATGATCTCCCTCGAACCCCTTCTCTCAC 302
Db 122 CC 63
QY 303 ACCCTCAGCAGCGCGCTTCTGCCAATCGACATCGCGCGCGCGCGCGCCACACCC 362
Db 62 CC 3

RESULT 12

ID ABQ52497 standard; DNA; 600 BP

AC ABQ52497

DT 12-JUL-2002 (first entry)
 XX

DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 39088
XY	

KM Human; cytosine methylation; 5-CpG-3'; uracil; cytosine; diagnosis;
 KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism
 KM SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2
 YY

PD 07-MAR-2002

XX 01-SEP-2001; 2001WO-EP010074

PR	01-SEP-2000; 2000DE-01043826.
PR	01-SEP-2000; 2000DE-01044543.

XX (EBTG-) EPITGENOMICS AG
PA

Oliver A. Dienerhacker C

XX
DB WBT. 2002-371829/40

Nettermining the deat

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridized to two classes, each with at least one member
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridization to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridized to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. AB033410-
CC AB034121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention

Sequence 600 BP; 44 A; 478 C; 51 G; 26 T; 0 U; 1 Other;

Query Match	18.1%;	Score	94.6;	DB	6;	Length	600;
Best Local Similarity	56.6%;	Pred.	No. 9e-09;				
Matches	175;	Conservative	0;	Mismatches	134;	Indels	0;
						Gaps	0

[illegible]

RESULT 13

ID ABQ52496 standard; DNA; 600 BP

QY 207 GGCCTGACATCCCCCTTGCTGGTGACCTTTTCATCTCAACCCCTCTCCACAGCG 266

Db 488 TTCCCCNTCCCCNTCCCTCTNTNCNCCCCCCCCCCCCCTCTCTTNTCTCTC 547

QY 267 ACCACTAGCTCCCTCTCTGACCCCTTCTCTCTCAACCTCAACGACCGAGTCTCCAA 328

Db 548 CCCCCCCCCCTCTCTCTCCCCCCCTCCCGCNCCTCTTCCCTCTCNCCTCCCCCNC 607

QY 327 CTGAGAGTGGCCCTTGCGCGCGCGCGGCCCAACCCAGAGCCACAGAGAGTGAAC 386

Db 608 CCCCCCCCCCTCTCNCNCCCCCCCCCAACCCCCCCCCCTTNNCNCNCCCTCNC 667

QY 387 TATGTCAAGAGTCAACCGAGCGC 410

Db 668 CCACTTCCCAACCTCTCTCTCTC 691

RESULT 15

ID	ABT09678 standard; DNA; 1064 BP.
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

AC ABT09678;

DT 02-DEC-2002 (first entry)

Human PAL-18 polynucleotide SEQ ID NO: 29.

Human, PAL-18; cancer; chromosome 1q41; prostate cancer; colon cancer
breast cancer; cyclostatic; gene; ds.

OS Homo sapiens.

PN US2002106765-A1.

PD 08-AUG-2002.

PF 12-MAR-2001; 2001US-00804682.

PR 10-MAR-2000; 2000US-0188586P.

PA (KIND/) KINDERS R J.

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Conclusion**
 6. **References**

PI Kinders RJ, Corey MJ;

DR WPI; 2002-697869/75.

PT New isolated PAL-18 polypeptide, useful for diagnosing, characterizing PT and treating disease and in determining disease susceptibility.

PS Claim 1; Page 47-48; 150pp; English.

CC The present invention relates to human PAL-18 polypeptides and
CC polynucleotides. The PAL-18 gene is found on chromosome 19q1. The
CC sequences can be used to diagnose, monitor and treat cancers,
CC particularly breast, colon and prostate cancers. The present sequence is
CC a PAL-18 polynucleotide shown in the invention

SQ Sequence 1064 BP; 93 A; 601 C; 51 G; 194 T; 0 U; 125 Other;

Query Match	17.6%	Score 92.2;	DB 6;	Length 1054;
Best Local Similarity	49.7%;	Pred. No. 2.5e+08;		
Matches 169; Conservative	0;	Mismatches 171;	Indels 0;	Gaps 0

QY	32	CACGCCACATCCAGAGGACACTGAGCAGAGAGAGTCCACACGGCGCGGACCATGATGTC	91
Db	472	CCCCCCCCCCCCGNNCCCTGAGNCTGCTC	533
QY	92	TCACACACGTTAGCGCCACCTCTCGACGACGCCCTTCTCCGCTTCCACGCGGCTCTC	153
Db	532	CCCCCTCTCTACCCCCCTCCNCCCNCCNCCNCCCAAGCCCCCCCCCTCTCTCCCTCT	593
QY	152	CTCCTCTGCCCCCTCTCTCTCCACACCGCGCGCTTCTCTCACCTGACCATCTTTTGCCCT	213

Db 592 CCGTCCCGCTCCCCCGCTCGNCGCCCGCCCGCCCGCTCCCTCCGCCCCCCCCCCCC 651

Qy 212 CGACTCCCCCTTGGCTGCTGACCTTTGCAATCTCAGCCCTTCTCCCGCAGCGCGACAC 2711

Db 652 NCGCCCCCTCCCTCCCCCCCCCTCCGCGCCGCTTCTGCGCCCGCCCGCCCTTACC 7111

Qy 272 TAGCTCCCTCTCGACCCCTTCTCTCTCGACACCTCAGCGCGCTTCTCCACTCGA 3311

Db 712 TNCNCGCCNCTGNCCTGNCNCGCCGCGCCGCTTCCCCCTTACCGCCCGCTTCCCG 7711

Qy 332 GCTGGCGCTCGCGCGCGCGCGCGACGACCGAGGCGACG 371

Db 772 TCCCCCCCCGNCNCGCCGCTTCCCCCCCCCGCCGCGCCCGCC 811

```
Search completed: April 14, 2004, 16:15:52
Job time : 438 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 14:27:16 ; Search time 97 Seconds
(without alignments)
2992.158 Million cell updates/sec

Title: US-09-669-817A-4

Perfect score: 523
Sequence: 1 accatcatcagcagatcacca.....gaccgcaagtcggaagtcgga 523

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109.2	20.9	7218	1	US-08-232-463-14 Sequence 14, Appl
2	98	18.7	320	3	US-09-165-264-14 Sequence 14, Appl
3	97.4	18.6	320	3	US-09-165-264-13 Sequence 13, Appl
4	97	18.5	319	3	US-09-165-264-8 Sequence 8, Appl
5	95.8	18.3	320	3	US-09-165-264-7 Sequence 7, Appl
6	95.4	18.2	318	3	US-09-165-264-12 Sequence 12, Appl
7	95.2	18.2	320	3	US-09-165-264-11 Sequence 11, Appl
8	81.2	15.5	1996	2	US-09-249-585A-4 Sequence 4, Appl
9	81.2	15.5	1991	2	US-09-130-114-2 Sequence 2, Appl
10	78	14.9	53526	3	US-08-658-136-2 Sequence 2, Appl
11	78	14.9	53577	3	US-08-658-136-1 Sequence 1, Appl
12	76.6	14.6	4466	4	US-09-410-551B-20 Sequence 20, Appl
13	76.6	14.6	4478	4	US-09-410-551B-16 Sequence 16, Appl
14	76.6	14.6	4547	4	US-09-410-551B-22 Sequence 22, Appl
15	76.6	14.6	4571	4	US-09-410-551B-18 Sequence 18, Appl
16	76.6	14.6	77536	4	US-09-410-551B-1 Sequence 1, Appl
17	76.4	14.6	12001	1	US-08-458-568A-11 Sequence 11, Appl
18	74.4	14.2	15231	3	US-09-128-155-16 Sequence 16, Appl
19	73.8	14.1	1996	4	US-09-249-585A-2 Sequence 2, Appl
20	73.8	14.1	1996	4	US-09-410-399-3 Sequence 3, Appl
21	73.8	14.1	2580	4	US-09-050-863-2 Sequence 2, Appl
22	73.8	14.1	2580	4	US-09-359-081-2 Sequence 2, Appl
23	73.8	14.1	5452	4	US-09-130-114-1 Sequence 1, Appl
24	73.8	14.1	8705	4	US-09-647-344A-14 Sequence 14, Appl
25	73.8	14.1	9600	4	US-08-910-647-1 Sequence 1, Appl
26	73.8	14.1	9600	4	US-09-620-925-1 Sequence 1, Appl
27	73.8	14.1	10596	1	US-07-884-811-15 Sequence 15, Appl

ALIGNMENTS

28	73.8	14.1	10596	1	US-07-885-971-15 Sequence 15, Appl
29	73.8	14.1	10596	1	US-08-087-783A-15 Sequence 15, Appl
30	73.8	14.1	10596	1	US-08-194-088B-15 Sequence 15, Appl
31	73.8	14.1	10596	2	US-08-194-087-15 Sequence 15, Appl
32	73.8	14.1	10596	5	PCT-US93-04648-15 Sequence 15, Appl
33	73.8	14.1	16080	4	US-03-724-566A-48 Sequence 48, Appl
34	73.6	14.1	77536	4	US-09-410-551B-1 Sequence 1, Appl
35	72.6	13.9	4897	6	5196516-7 Patent No. 5196516
36	71.4	13.7	51259	3	US-08-781-891-209 Sequence 209, App
37	71.4	13.7	51259	4	US-09-618-166-209 Sequence 209, App
38	65.4	12.5	925	3	US-08-858-003-1 Sequence 1, Appl
39	65.4	12.5	925	3	US-09-078-166-1 Sequence 1, Appl
40	65.4	12.5	925	3	US-08-997-467-1 Sequence 1, Appl
41	64.4	12.3	438	4	US-09-252-991A-8422 Sequence 8422, Ap
42	64.4	12.3	474	4	US-09-252-991A-8365 Sequence 8365, Ap
43	64.4	12.3	1761	4	US-09-252-991A-8423 Sequence 8423, Ap
44	63.8	12.2	31571	1	US-08-323-443B-1 Sequence 1, Appl
45	63.2	12.1	4403765	3	US-09-103-840A-2 Sequence 2, Appl

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpC-F15
US-08-232-463-14
Query Match 20.9%; Score 109.2; DB 1; Length 7218;

Best Local Similarity 3.2%; Pred. No. 8.8e-14;
Matches 12; Conservative 260; Mismatches 98; Indels 0; Gaps 0;

```
QY 23 TCTCTTCACACTCCCACTCCAGGAGAGCTGAGAGAGCTCCACCGCGCGGAGCAC 82
DB 1085 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1144
QY 83 CATGAGTTCTACACGCTTACGCCACCTCTCGACGAGCCCTTCTCCCTCCAC 142
DB 1145 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1204
QY 143 GCGCGCTCTCTCTCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 202
DB 1205 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1264
QY 203 CTTCGAGCTGAGCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 262
DB 1265 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1324
QY 263 GCGGACACAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 322
DB 1325 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1384
QY 323 CCAACTGAGCTGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 382
DB 1385 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1444
QY 383 CACCTATGTC 392
DB 1445 CTTCTATCTC 1454
```

RESULT 2

US-09-165-264-14/c
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

Query Match 18.7%; Score 98; DB 3; Length 320;
Best Local Similarity 57.0%; Pred. No. 1.1e-11;
Matches 179; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
QY 49 CAGCTGAGCAGAGCTGACCGCGCGCGGACATGAGTTCTACACACGTTAGCGCC 108
DB 317 CTGCTGATGATGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 258
QY 109 ACCCTCTGAGAGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 168
DB 257 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 198
QY 169 TCTCCACCGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 228
DB 197 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 138
QY 229 CTGACCTTTTCATTCACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAC 288
DB 137 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 78
QY 289 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 348
```

```
DB 77 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 18
QY 349 GGGCCCCACACCC 362
DB 17 CCCCCCCCCCCCCC 4
```

RESULT 3

US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

Query Match 18.6%; Score 97.4; DB 3; Length 320;
Best Local Similarity 57.3%; Pred. No. 1.5e-11;
Matches 176; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
QY 65 TCCACCGCGCGCGGACACATGAGTTCTACACACCGTTACGCCCACTCTCTGAGAGCC 124
DB 307 TCCAGTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 248
QY 125 CTTCTCTCCCTTCCACGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 184
DB 247 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 188
QY 185 TTTCTTCACATGCGCATTTTGGCTTGAATCTCTCTCTCTCTCTCTCTCTCTCT 244
DB 187 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 128
QY 245 CCACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 304
DB 127 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 68
QY 305 CCTCACGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 364
DB 67 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 8
QY 365 GCCCAGC 371
DB 7 CCCCCCC 1
```

RESULT 4

US-09-165-264-8/c
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence

APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 53526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 14.9%; Score 78; DB 3; Length 53526;
Best Local Similarity 51.4%; Pred. No. 2.2e-07;
Matches 180; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 26 CCTTCACACTCCCACTCCAGCGAGCTGAGAGAGCTCCAGCGCGCGGCGGCGACCAT 85
DB 34410 CTTTCCCTCCTCCTCCTCCGCGTGAAGCCCTCCACCTGTCCTCCAGCCCTCCTCCT 34469
QY 86 GAGTTCTCAGCAGCGTTAGCGCCACCTCTCTGAGAGACCCCTTCTTCCCTTCCAGCGCC 145
DB 34470 AGCCCTCCCTCCCTCCCTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCTCTCTCT 34529
QY 146 GTCTCTCTCTCTCTGCGCCCTTCTCTCTCCAGCGCGGCTTCTCTCACTGCGCATCTT 205
DB 34530 CCCCCTCCTCCTCCTCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 34589
QY 206 CGCCCTGAGTCCCGCTCTGCTGCTGACCTTTCATCTGACACCCCTTCTCTCCAGCGCC 265
DB 34590 CCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 34649
QY 266 GACCACTAGTCCCTCCTCTGAGACCCCTTCTCTCTCCTCCTCCTCCTCCTCCTCCTCCT 325
DB 34650 CCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 34709
QY 326 ACTGAGCTGCGCCCTGCG 375
DB 34710 CTTCCCT 34759

RESULT 11
US-08-658-136-1
Sequence 1, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LINDER, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERIINO, GREGORY
APPLICANT: QIAN, PENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 53577 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

Query Match 14.9%; Score 78; DB 3; Length 53577;
Best Local Similarity 51.4%; Pred. No. 2.2e-07;
Matches 180; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 26 CCTTCACACTCCCACTCCAGCGAGCTGAGAGAGCTCCAGCGCGCGGCGGCGACCAT 85
DB 34405 CTTTCCCTCCTCCTCCTCCGCGTGAAGCCCTCCACCTGTCCTCCAGCCCTCCTCCT 34464
QY 86 GAGTTCTCAGCAGCGTTAGCGCCACCTCTCTGAGAGACCCCTTCTTCCCTTCCAGCGCC 145
DB 34465 AGCCCTCCCTCCCTCCCT 34524
QY 146 GTCTCTCTCTCTCTGCGCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 205
DB 34525 CCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 34584
QY 206 CGCCCTGAGTCCCGCTCTGCTGCTGACCTTTCATCTGACACCCCTTCTCTCCAGCGCC 265
DB 34585 CCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 34644
QY 266 GACCACTAGTCCCTCCTCTGAGACCCCTTCTCTCTCCTCCTCCTCCTCCTCCTCCTCCT 325
DB 34645 CCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 34704
QY 326 ACTGAGCTGCGCCCTGCG 375
DB 34705 CTTCCCT 34754

RESULT 12
US-09-410-551B-20
Sequence 20, Application US/09410551B
Patent No. 6503737
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.
APPLICANT: REEVES, CHRISTOPHER
APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SANTU, DANIEL
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
TITLE OF INVENTION: CONSTRUCTS THEREFOR
FILE REFERENCE: 30062-20026.00
CURRENT APPLICATION NUMBER: US/09/410,551B
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ ID NOS: 72

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 20

LENGTH: 4466

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic

OTHER INFORMATION: PKS synthase fragment

FEATURE:

NAME/KEY: CDS

LOCATION: (9)..(4454)

US-09-410-551B-20

Query Match 14.6%; Score 76.6; DB 4; Length 4466;

Best Local Similarity 53.4%; Pred. No. 3.2e-07;

Matches 183; Conservative 0; Mismatches 159; Indels 1; Gaps 1;

32 CACTCCCACTCCAGGAGCTGAGCAGAGAGCTCCACCGCGCGGCACCATGAGTTC 91

3741 CACCCCAACACCCCTCTCCCTGCGCCCACTGCGCCACCTCGACACCCCACTCCGC 3800

92 TCACCACTGTTAGCGCCCACTCTCTGACGACCCCTTCTCCCTCCGACGCGCTTC 151

3801 -CTCACCACTGACACCTCTCCACACCCCACTCACTCCCTCCGACACACCCCACT 3859

152 CTCTCTGCGCCCT 211

3860 CACCAACACCCCTCTCAACCCGAAACAGCATATATATATATATATATATATATAT 3919

212 CACTCCCT 271

3920 CGCCGGATCT 3979

272 TAGCT 331

3980 ACCCCCGACGCGACCCCGCGACCCGACCTCTCTCTCTCTCTCTCTCTCTCTCT 4039

332 GCTCGCCCT 374

4040 CGCACCACTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 4082

RESULT 13
US-09-410-551B-16

Sequence 16, Application US/09410551B

Patent No. 6503737

GENERAL INFORMATION:

APPLICANT: KOSAN BIOSCIENCES, Inc.

APPLICANT: REEVES, CHRISTOPHER

APPLICANT: CHU, DANIEL

APPLICANT: KHOSLA, CHAITAN

APPLICANT: SANTI, DANIEL

APPLICANT: MU, KAI

TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

TITLE OF INVENTION: CONSTRUCTS THEREFOR

FILE REFERENCE: 30062-20026.00

CURRENT APPLICATION NUMBER: US/09/410.551B

PRIOR FILING DATE: 1999-10-01

PRIOR APPLICATION NUMBER: US 60/139,650

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/123,810

PRIOR FILING DATE: 1999-03-11

PRIOR APPLICATION NUMBER: US 60/102,748

PRIOR FILING DATE: 1998-10-02

NUMBER OF SEQ ID NOS: 72

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16

LENGTH: 4478

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic

OTHER INFORMATION: PKS synthase fragment

FEATURE:

NAME/KEY: CDS

LOCATION: (3)..(4466)

US-09-410-551B-16

Query Match

Best Local Similarity 14.6%; Score 76.6; DB 4; Length 4478;

Matches 183; Conservative 0; Mismatches 159; Indels 1; Gaps 1;

32 CACTCCCACTCCAGGAGCTGAGCAGAGAGCTCCACCGCGCGGCACCATGAGTTC 91

3753 CACCCCAACACCCCTCTCCCTGCGCCCACTGCGCCACCTCGACACCCCACTCCGC 3812

92 TCACCACTGTTAGCGCCCACTCTCTGACGACCCCTTCTCCCTCCGACGCGCTTC 151

3813 -CTCACCACTGACACCTCTCCACACCCCACTCACTCCCTCCGACACACCCCACT 3871

152 CTCTCTGCGCCCT 211

3872 CACCAACACCCCTCTCAACCCGAAACAGCATATATATATATATATATATATAT 3931

212 CACTCCCT 271

3932 CGCCGGATCT 3991

272 TAGCT 331

3992 ACCCCCGACGCGACCCCGCGACCCGACCTCTCTCTCTCTCTCTCTCTCTCTCT 4051

332 GCTCGCCCT 374

4052 CGCACCACTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 4094

RESULT 14

US-09-410-551B-22

Sequence 22, Application US/09410551B

Patent No. 6503737

GENERAL INFORMATION:

APPLICANT: KOSAN BIOSCIENCES, Inc.

APPLICANT: REEVES, CHRISTOPHER

APPLICANT: CHU, DANIEL

APPLICANT: KHOSLA, CHAITAN

APPLICANT: SANTI, DANIEL

APPLICANT: MU, KAI

TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

TITLE OF INVENTION: CONSTRUCTS THEREFOR

FILE REFERENCE: 30062-20026.00

CURRENT APPLICATION NUMBER: US/09/410.551B

PRIOR FILING DATE: 1999-10-01

PRIOR APPLICATION NUMBER: US 60/139,650

PRIOR FILING DATE: 1999-06-17

PRIOR APPLICATION NUMBER: US 60/123,810

PRIOR FILING DATE: 1999-03-11

PRIOR APPLICATION NUMBER: US 60/102,748

PRIOR FILING DATE: 1998-10-02

NUMBER OF SEQ ID NOS: 72

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 22

LENGTH: 4547

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic

OTHER INFORMATION: PKS synthase fragment

FEATURE:

NAME/KEY: CDS

LOCATION: (9)..(4535)

US-09-410-551B-22

Query Match

Best Local Similarity 14.6%; Score 76.6; DB 4; Length 4547;

Matches 183; Conservative 0; Mismatches 159; Indels 1; Gaps 1;

QY 32 CACTCCACTCCAGCGGAGCTGAGAGAGCTCCACCGCGCGCGCCACCATGATGTTG 91
DB 3822 CACCCCAACACCCCGCTCCCGCTGAGCCCAACTCGCACCCCTGAGACCCCAACCTCCGCG 3881
QY 92 TCACACACGTTAGGCGCACTCTCTGACAGAACCCCTTCTTCCCTTCCACCGCGCTCTG 151
DB 3882 -CTCACACACACACACCCCTCCACACACCCCAACCTCCACCCCTCCACACACACACCCCAAC 3940
QY 152 CTCCTCTGCGCCCTTCTCTCTCTCCCGACCGCGGTTCTCTCAACCTGCGCATTTCTGCGCT 211
DB 3941 CACCAACACCCCGCTCAACCCCGAACAACGCGATCATATCATCGGAGGCTCCGCACTT 4000
QY 212 CGACTCCCGCTTGGCTGCTGACCTTTCATCTGACACCCCTTCTCTCCGACGCGACAC 271
DB 4001 CGCGGAGATCTGCGCGCGCGACCTGAAACACACCCCAACCTTCTCTCTCCGCGACCC 4060
QY 272 TAGCTCCCTCTGAGACCCCTTCTCTCTCAACACCTCCACCGAGCGGTCTCCCACTGGA 331
DB 4061 ACCCGCGAGCGCACCCCGCGACCACTCTCTCTGAGCTGCGGACCCCGACCACT 4120
QY 332 GCTGCGCTTGGCG 374
DB 4121 CGCGACACACCTCAACCAACATCCCGCAACCCCTCAACCGCATC 4163

RESULT 15
US-09-410-551B-18

; Sequence 18, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: KU, KAI
; TITLE OF INVENTION: POLYMERASE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 4571
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9) ..(4559)
US-09-410-551B-18

Query Match 14.6%; Score 76.6; DB 4; Length 4571;

Best Local Similarity 53.4%; Pred. No. 3.3e-07;

Matches 183; Conservative 0; Mismatches 159; Indels 1; Gaps 1;

QY 32 CACTCCACTCCAGCGGAGCTGAGAGAGCTCCACCGCGCGCGCCACCATGATGTTG 91
DB 3846 CACCCCAACACCCCGCTCCCGCTGAGCCCAACTCGCACCCCTGAGACCCCAACCTCCGCG 3905
QY 92 TCACACACGTTAGGCGCACTCTCTGACAGAACCCCTTCTTCCCTTCCACCGCGCTCTG 151
DB 3906 -CTCACACACACACACCCCTCCACACACCCCAACCTTCAACCCCTTCAACACACACCCCAAC 3964

QY 152 CTCCTCTGCGCCCTTCTCTCTCTCCCGACCGCGGTTCTCTCAACCTGCGCATTTCTGCGCT 211
DB 3965 CACCAACACCCCGCTCAACCCCGAACAACGCGATCATATCATCGGAGGCTCCGCACTT 4024
QY 212 CGACTCCCGCTTGGCTGCTGACCTTTCATCTGACACCCCTTCTCTCCGACGCGACAC 271
DB 4025 CGCGGAGATCTGCGCGCGCGCGACCTGAAACACACCCCAACCTTCTCTCTCCGCGACCC 4084
QY 272 TAGCTCCCTCTGAGACCCCTTCTCTCTCAACACCTCCACCGAGCGGTCTCCCACTGGA 331
DB 4085 ACCCGCGAGCGCACCCCGCGACCACTCTCTCTGAGCTGCGGACCCCGACCACT 4144
QY 332 GCTGCGCTTGGCG 374
DB 4145 CGCGACACACCTCAACCAACATCCCGCAACCCCTCAACCGCATC 4187

Search completed: April 14, 2004, 16:19:59
Job time : 104 secs

Db 164 GCTTCTCTCTTGAGAGACCAACCCTTCTTACCCAACTCTCTCCGCTGGAGATACCTCC 223

QY 193 CTTGCCATTTCTTGCCCCCTGACTCCCCCTTGGCTGACCTTTTCATCTCAACCCT 252

Db 224 GCTCCCGATCGAGACACCTTCGACTTCACCTCTCTCCCTCCCTCCACAGCCG 283

QY 253 TCTCTCCACAGCCGAC---CACTAGCTCCCTCTCGACCCCTTCTCTCTCCACACCTCA 309

Db 284 CCGTCTCGCGCCCCCGCGCGCGGCTTGCACCTTCTCTCTCTGACACCTCG 343

QY 310 CCGACCGGCTGTCCCAACTCGAGCTCGCCCTCGCGCGCCCGCCACACCCAGGCCCCA 369

Db 344 GCCACCCGATTCCTGCTCTCTCGAGCGCGCGCTCG-----CCTGCCCCAC 388

QY 370 CCACACAGAAATGACACCTATGTACCGAGTCAACCGGCGCGATGTCAAGTGAACACG 429

Db 389 CCGGCGCAATTAATCTACGCGCCGCCGAGGCCAACGGCAGAACTCAAGTGAACCGCG 448

QY 430 AGGACAAAGCTCGCGCGCGCGAGAGGAGGTGCTCAAGTGGAGAGCTGAATCTCCCT 489

Db 449 AGGACAGGCCC---CGCGGGGCGCGCAACCTCAAGTGGAGGCGGAGCTCAGACCCGCG 505

QY 490 ACGACGATGATTTGACCGCGCAAGTGAAGTGGGA 523

Db 506 ACACCGACGCTTGCACCGCAATGGAATGGGA 539

```

RESULT 2
US-10-424-599-20495
: Sequence 20495, Application US/10424599
: Publication NO. US20040031072A1
: GENERAL INFORMATION:
: APPLICANT: La Rosa Thomas J
: APPLICANT: Kovalic David K
: APPLICANT: Zhou Yihua
: APPLICANT: Cao Yongwei
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53223)B
: CURRENT APPLICATION NUMBER: US/10/424,599
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 285684
: SEQ ID NO 20495
: LENGTH: 815
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(8-5)
: OTHER INFORMATION: unsure at all n locations
: FEATURE:
: OTHER INFORMATION: Clone ID: PAT_MRT3847_118511C.1
US-10-424-599-20495

```

Query Match	20.2%	Score 105.4	DB 12	Length 815
Best Local Similarity	55.3%	Pred. No. 1,le-16		
Matches	205	Conservative	0	Mismatches 166
			Indels	0
			Gaps	0

QY	1	AATGATACGACGATACCAAGTCTCTTCACACTCCACTCAGCGGAGGTAGACAGA	60
Db	280	ACCCGACCCCTATCCCAAGCCACACACCCCCCGGCTCCCTCCCTCCCT	339
QY	61	GAGCTCACCGCGCGCGGCACCATGATTTCACACCGTTAGCGCCACCTCTTGAG	120
Db	340	CCCCCCCCCCCCCCCCCCCCCGACCCCCCGCGCCCCCCCCCCCCCCCCCCCC	399
QY	121	AACCTTCTTCCCTTCCACAGCGCGTCTCTCTCCCTCGGCCCTTCTCTCCACAGG	180
Db	400	CC	459
QY	181	CCGTTTCTTCACCTGCGCATTTCTTGCGCTTGACTCCCCCTTGCGTGTGACCTTTC	240

Db	460	CC	512
Qy	241	ATTTCAACCCCTCTCTCCGCAGCGGCACTAGTCTCTCTGACCTTCTCTCC	300
Db	520	CC	579
Qy	301	ACACCTTCAACGAGCGGCTGTCGCAATCGAGCTCGAGCTCGAGCGCGGCGCCACAC	360
Db	580	CC	639
Qy	361	CGAGGCCACG	371
Db	640	CCCCCCCCCCCC	650

```

RESULT 3
US-10-424-599-107958/c
; Sequence 107958, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalick David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 107958
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(771)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT1847_68501C.1
US-10-424-599-107958

```

Query	Match	Similarity	19.5%	Score 102	DB 12	Length 771
Beet	Local	Similarity	53.3%	Pred. No. 7.6e-16		
Matches	186	Conservative	0	Mismatches	163	Indels
						Gaps
						0
QY	26	CCTTCACACTCCCACTCCAGCGGAGCTGACGACAGAGACTCCACCGCGCGGACCAAT	85			
DB	753	CCCA	694			
QY	86	GAGTTTACCAACGGTTACGGCCACACTCTCTGAGAGACCCCTTTCTTCCCTCCACGGC	145			
DB	653	CCCCCCCCCCCCCCCCCTTNNNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	634			
QY	146	GTCCTCTCTCTGACGCCCCCTTCTCTCTCCCAACCGACGTCCTTCTCCACCTGGCCATTCT	205			
DB	633	CCCCCCCCCCCCCCCCCCCCCCCCCAACCCACCCGACCCGCCCCCCCCNNNNCCCCCCCC	574			
QY	206	CGCCTTGACACTCCCTCTTGGCTGCTGACCCCTTTCATCTCCACACCCCTTCTCCCAAGC	265			
DB	573	CCC	514			
QY	266	GACCACTAGTCCCTCCCTCTGACCCCTTCTCTCTCTCCACACCTCTGACCGAGGTCTCCA	325			
DB	513	CCC	454			
QY	366	ACTGAGTGGCCTCGCGCGCCGCGCCCAACCCAGGCCACAGC	374			
DB	453	CCCCCCCCCCCCCCCCCCCCCCCCCAACCCCCCCCCCCCCCCCCCCCAACCG	405			

Sequence 47, Application US/10032393
Publication No. US20030027286A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Wall, Daniel
APPLICANT: Gross, Molly
TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
FILE REFERENCE: ELITRA.010A
CURRENT APPLICATION NUMBER: US/10/032,393
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/259,434
PRIOR FILING DATE: 2000-12-27
PRIOR APPLICATION NUMBER: 09/948,993
PRIOR FILING DATE: 2001-09-06
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
NUMBER OF SEQ ID NOS: 68
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 47
LENGTH: 12733
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Vector pBPEF14
US-10-032-393-47

Query Match 19.5%; Score 102; DB 14; Length 12733;
Best Local Similarity 54.9%; Pred. No. 6.3e-16;
Matches 201; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 6 ATGAGCAGATCACAAGTCTCTTCCACTCCCACTCCAGCGGAGTGAAGAGACT 65
DB 5357 ATGAGTGAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5298
QY 66 CCAACGCGCGCGGCAACATGAGTTCTACACAGCTTACGCCACTCTCTGACAGCC 125
DB 5297 CC 5238
QY 126 TTCTTCCCTTCCGACGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
DB 5237 CC 5178
QY 186 TCTTCACCTGCGCATTTCTTGCGCTGAGTCCCTTCTGAGCTTTCATCTC 245
DB 5177 CC 5118
QY 246 CACCCCTTCTCCGACGCGGCAACTAGCTCCCTCTGAGCCCTTCTCTCACACC 305
DB 5117 CC 5058
QY 306 CTCACGACCGGCTCTCCCACTGAGCTGCGCGCGCGGCGCCACACCCCGAG 365
DB 5057 CC 4998
QY 366 CCCACC 371
DB 4997 CCCCCC 4992

RESULT 5
US-10-032-393-8/c
Sequence 8, Application US/10032393
Publication No. US20030027286A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Wall, Daniel
APPLICANT: Gross, Molly
TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
FILE REFERENCE: ELITRA.010A
CURRENT APPLICATION NUMBER: US/10/032,393
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/259,434
PRIOR FILING DATE: 2000-12-27

PRIOR APPLICATION NUMBER: 09/948,993
PRIOR FILING DATE: 2001-09-06
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
NUMBER OF SEQ ID NOS: 68
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 12739
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Vector pBPEF1
US-10-032-393-8

Query Match 19.5%; Score 102; DB 14; Length 12739;
Best Local Similarity 54.9%; Pred. No. 6.3e-16;
Matches 201; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 6 ATGAGCAGATCACAAGTCTCTTCCACTCCCACTCCAGCGGAGTGAAGAGACT 65
DB 5363 ATGAGTGAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5304
QY 66 CCAACGCGCGCGGCAACATGAGTTCTACACAGCTTACGCCACTCTCTGACAGCC 125
DB 5303 CC 5244
QY 126 TTCTTCCCTTCCGACGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
DB 5243 CC 5184
QY 186 TCTTCACCTGCGCATTTCTTGCGCTGAGTCCCTTCTGAGCTTTCATCTC 245
DB 5183 CC 5124
QY 246 CACCCCTTCTCCGACGCGGCAACTAGCTCTCTCTCTCTCTCTCTCTCT 305
DB 5123 CC 5064
QY 306 CTCACGACCGGCTCTCCCACTGAGCTGCGCGCGCGGCGCCACACCCCGAG 365
DB 5063 CC 5004
QY 366 CCCACC 371
DB 5003 CCCCCC 4998

RESULT 6
US-10-017-161-1857/c
Sequence 1857, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASHI, KIYOSHI
APPLICANT: AKIYAMA, YUTIKA
APPLICANT: ABURAYANT, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1857
LENGTH: 3163
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: source
LOCATION: (1)..(3163)
FEATURE:
NAME/KEY: CDS

LOCATION: (201)..(2963)
FEATURE:
NAME/KEY: modified_base
LOCATION: (556)..(655)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (658)..(669)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (671)..(697)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (704)..(707)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (709)..(736)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (739)..(743)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (745)..(747)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (750)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (752)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (754)..(756)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (761)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (763)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (818)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (843)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (897)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (920)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (941)
OTHER INFORMATION: a, t, c, g, unknown or other

FEATURE:
NAME/KEY: modified_base
LOCATION: (993)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (995)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1066)..(1067)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1082)..(1084)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1105)..(1107)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1130)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1147)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1162)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1202)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1233)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1247)..(1248)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1265)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1485)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1500)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1544)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1601)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1605)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1635)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:

```

1 TITLE OF INVENTION: GARGANSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
2 FILE REFERENCE: 084333/116
3 CURRENT APPLICATION NUMBER: US/10/292,798
4 PRIOR FILING DATE: 2002-11-23
5 PRIOR APPLICATION NUMBER: 10/017,161
6 PRIOR FILING DATE: 2001-12-18
7 PRIOR APPLICATION NUMBER: JP 2001-246789
8 PRIOR FILING DATE: 2001-06-18
9 NUMBER OF SEQ ID NOS: 2070
10 SOFTWARE: PatentIn Ver. 2.1
11 SEQ ID NO 1513
12 LENGTH: 3163
13 TYPE: DNA
14 ORGANISM: Homo sapiens
15 FEATURE:
16 LOCATION: source
17 FEATURE:
18 LOCATION: (1)..(3163)
19 FEATURE:
20 NAME/KEY: CDS
21 LOCATION: (201)..(2963)
22 FEATURE:
23 NAME/KEY: modified_base
24 LOCATION: (356)..(655)
25 OTHER INFORMATION: a, t, c, g, unknown or other
26 FEATURE:
27 NAME/KEY: modified_base
28 LOCATION: (658)..(669)
29 OTHER INFORMATION: a, t, c, g, unknown or other
30 FEATURE:
31 NAME/KEY: modified_base
32 LOCATION: (709)..(736)
33 OTHER INFORMATION: a, t, c, g, unknown or other
34 FEATURE:
35 NAME/KEY: modified_base
36 LOCATION: (739)..(747)
37 OTHER INFORMATION: a, t, c, g, unknown or other
38 FEATURE:
39 NAME/KEY: modified_base
40 LOCATION: (750)..(750)
41 OTHER INFORMATION: a, t, c, g, unknown or other
42 FEATURE:
43 NAME/KEY: modified_base
44 LOCATION: (752)..(752)
45 OTHER INFORMATION: a, t, c, g, unknown or other
46 FEATURE:
47 NAME/KEY: modified_base
48 LOCATION: (761)..(761)
49 OTHER INFORMATION: a, t, c, g, unknown or other
50 FEATURE:
51 NAME/KEY: modified_base
52 LOCATION: (763)..(763)
53 OTHER INFORMATION: a, t, c, g, unknown or other
54 FEATURE:
55 NAME/KEY: modified_base
56 LOCATION: (818)..(818)
57 OTHER INFORMATION: a, t, c, g, unknown or other

```

FEATURE:	NAME/KEY: modified base	LOCATION: (943) ..(943)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (892) ..(892)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (897) ..(897)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (920) ..(920)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (941) ..(941)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (993) ..(993)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (995) ..(995)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1066) ..(1067)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1082) ..(1084)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1105) ..(1107)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1162) ..(1162)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1247) ..(1248)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1261) ..(1261)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1485) ..(1485)	a, t, c, g, unknown or other
FEATURE:	NAME/KEY: modified base	LOCATION: (1485) ..(1485)	a, t, c, g, unknown or other

NAME/KEY: modified_base	
LOCATION: (1500) .. (1500)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1544) .. (1544)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1601) .. (1601)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1605) .. (1605)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1635) .. (1635)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1637) .. (1637)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1654) .. (1654)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1674) .. (1674)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1695) .. (1695)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1852) .. (1852)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1866) .. (1866)	
OTHER INFORMATION: a, t, c, g, unknown or other	
FEATURE:	
NAME/KEY: modified_base	
LOCATION: (1968) .. (1980)	
OTHER INFORMATION: a, t, c, g, unknown or other	

Query Match	19.4%;	Score 101.4;	DB 15;	Length 3163;
Best Local Similarity	51.1%;	Pred. No. 9.7e-16;		
Matches 207; Conservative	0;	Mismatches 198;	Indels 0;	Gaps 0;

QY ATGAGAGATATACCAAGTCTCTTTACACCTCCCACTCCAGGGAGGTGACAGAGAGCT 65
 1270 AATACCCCCNCCCGAACCCTCCNCCCTCCCTCCNCGGCCCCCGGGCCCCCCCCCCC 1211
 66 CCAACGCGCGCGGAGCACAATGATTCTACCAACGTTACGCGCACTCTTGACGACCCC 1255
 Db CCCCCCCNCGCGCC 1151
 QY 1210 CCCCCCCNCGCGCC 1151
 QY 126 TTCTTCCCTTCCCAACGCGCGTCTCTCTCTCTGACCCCTTCTCTGACCAACGCGCTT 1855
 Db 1150 CCGNCCCCCCCCCGCGCCCCCAGCCCAACCCCCCCCCCCCCCCCCCGCCNCCGACACCCCCC 1099
 QY 186 TCGTCACACTGCGCATTTCTTGCGCCTGAGCTCCCCCTTGCGTGTGACCTTTTCATCTC 2451
 Db 1090 CCCCCCNMNCCTCCCCCCCCCCCCCGCCNCCCTCCCCCCCCCCCCCCCCCCCCCCC 1030
 QY 246 CACCCCTTCTCTCCCAACGCGGACACATGAGCTCCCTCTCGACCCCTTCTCTCTCAACC 3055
 Db 1030 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCNCCGCCCCCCCCCCCCCCC 9711
 QY 306 CTCACGACCGAGTCTCCCAACTGAGCTGCGCCTCGCGCGCCGACCCCAACCCACG 3555


```

LOCATION: (298)..(300)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (302)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (307)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (314)..(315)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (334)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (339)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (343)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (345)..(348)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (351)..(352)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (357)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (359)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (362)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (365)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (367)..(368)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (370)..(373)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (376)..(378)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base

```

[illegible]

QY	124	CCCTCTTCCGCTTCCACGCGCGTCCCTCTCTCTGCTCCCTTCTCTCCGACCGGCG	183
Db	3106	CCGNNCCCCCCKNNNNCCCCNNCCCCNNCNCNCCCCCCCCCCKNNNNCCCCCCCCC	3047
QY	184	TTTCTCCACCTGCGCCATTCTTGCGCCTCTGACTCCCTCTGCTGTGACCTTTCCATC	243
Db	3046	NNCCCCCCCCCKNNNNCKNNCCCCCNCCCKNNCCCCCCCCCCCCCKNNCKNNCKNNNN	2987
QY	244	TTCAACCCCTTCTCTCCCGCGGACACTGAGCTCCTCTCTGACCCCTTCTCTCCACA	303
Db	2986	CCCCCCCCCCCCCCCCCCCCCKNNCKNNCCCCCCCCCKNNNNCCCCCKNNCCCCCKNN	2927
QY	304	CCCTCACGACCGGCTGTCCCACTGAGGTGGCGCTCGCGCGCGCGGCCCAACGCCA	363
Db	2926	NCCTCCGNNCCCCCNCCCCCKNNCCCKCCCCCCCCCKNNNNCCCCCKNNCCCCCN	2867
QY	364	GGCCACACGACGAGAGTGACCTATGTCAACGAGTCCACCGGCGGACGTCAAGTGA	423
Db	2866	NCGNNCKNNNNNNNNNNCKNNNNNNNNCCCCCCCCCCCCCKNNNNNNNNNNNNCC	2807
QY	424	CCACCGAGGACAGCCTCGCGCGCGGCGAGAGGGTGTCTCAAGTGGAGAGCTGCACT	483
Db	2806	CCGNNCKNNCKNNNNCKNNCKNNCKNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCN	2747
QY	484	CCCCCTACGAC	494
Db	2746	NCCTCCCCCNC	2736

```

RESULT 9
US-10-292-798-1189/c
Sequence 1189, Application US/10292798
Publication No. US20030235833A1
GENERAL INFORMATION:
APPLICANT: SUWA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABRATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 064335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1189
LENGTH: 5452
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: source
FEATURE:
LOCATION: (1)..(5452)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(5252)
FEATURE:
NAME/KEY: modified_base
LOCATION: (1)..(140)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (145)..(145)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (147)..(148)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base

```

```

LOCATION: (162)..(162)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (165)..(165)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (168)..(168)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (172)..(172)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (189)..(189)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (191)..(193)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (196)..(196)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (199)..(199)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (214)..(214)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (218)..(218)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (221)..(221)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (225)..(226)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (229)..(229)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (232)..(232)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (235)..(236)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (246)..(246)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (248)..(252)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (256)..(257)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (261)..(263)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (265)..(266)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (271)..(273)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (276)..(277)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (283)..(285)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (296)..(296)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (298)..(300)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (302)..(302)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (307)..(307)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (314)..(315)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (339)..(339)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (343)..(343)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (345)..(348)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (351)..(352)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (357)..(357)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (359)..(359)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (362)..(362)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (365)..(365)
OTHER INFORMATION: a, t, c, g, unknown or other

```


[illegible]

```

RESULT 12
US-10-424-599-2320
; Sequence 2320, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 2320
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(580)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT3847_10209C.1
US-10-424-599-2320

```

Query Match	18.1%;	Score 94.6;	DB 12;	Length 560;
Best Local Similarity	54.9%;	Pred. No. 5e-14;		
Matches 169; Conservative	0;	Mismatches 139;	Indels 0;	Gaps 0

QY	6	CTCCACGCGCGCGGACGACATGATTCTACACACCGTTAGGCCACCTCTTGAGACG	123
Db	267	CC	326
QY	124	CGTTCTTCCGCTTCCACACGCGGTCTCTCTCTGAGCCCTTCTCTCTCCACACCGCGG	183
Db	327	CCCCCCCCNNNNCTCCCCCCCCCCCCCCCCCCCCCTTCCCCCCNNCCCCCCCCCCCC	386
QY	184	TTTCTCTACACGTCGACATCTTGCGGCTTGACCTCCCCCTTGCGTGTGACCCCTTTGATC	243
Db	387	CCCCCCCCCCCCCCCCCCTCTCCGCCCCCCCCCCCCCCCCCTTCCCCCCCCCCCC	446
QY	244	TCCACCGCTTCTCTCCACGCGGACCACTAGTCCCTCTCTGACCCCTTCTCTTCCACA	303
Db	447	NNCCCCCCCCCCCCCCCCCCCCCCCCCTTCCCTTCCCCCCCCCCNNCAGCCCNCG	506
QY	304	CGCTCACGACGCGCGTCTCCCACTGAGACTTGCGCGCGCGGCCACACCCCA	363
Db	507	ACCGCCCGCCCCCCCCCCCCCGGCCGCCCCCCCCCGGCGCCGCCGCCGCC	566
QY	364	GGCGCCACC	371
Db	567	CCCGGCCG	574

RESULT 13

```

US-10-424-599-54839
; Sequence 54839, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 54839
; LENGTH: 778

```

```

? ORGANISM: Glycine max
? FEATURE:
? NAME/KEY: unsure
? LOCATION: (1)..(778)
? OTHER INFORMATION: unsure at all n locations
? FEATURE:
? OTHER INFORMATION: Clone ID: PAT.MR1847_20531C.1
? OS:10-424-599-54639

```

Query Match	18.0%;	Score 94;	DB 12;	Length 778;
Best Local Similarity	52.6%;	Pred. No. 6.9e-14;		
Matches 202;	Conservative	0;	Mismatches 182;	Indels 0;
				Gaps 0;

Oy	24	CTCTTTCAACATCCCACTCAAGCGGACAGTGAAGAGACTCAACCGCGCGGCGAAC	83
Db	206	CC	265
Oy	84	ATGAGTTCTACCAACCGTTAAGCCCACTCTCGACGACCCCTTTCCTTCCACCG	143
Db	266	CCCCGACCC	325
Oy	144	CGTCTCTCTCTCTCTGCCCCCTTCTCTCTCCGACCGCGGTTTCTTCACATGCCATTG	203
Db	326	CCCCCCCCCCCCCTTCTCCCCCCCCCCCCCCCCCCCCCAACGACAAACCCCCCCCC	385
Oy	204	TTGGCCCTGAGTCCGCCCTTGAGTGTGAGCCCTTTCATCCAGCCGTTTCCGCCACG	263
Db	386	CCCTCTCCCCCCCCCCCCCTTCTTCAACAACCTCCCCCCCCCCCCCTTCTCCCC	445
Oy	264	CGACCACTAGTCCCTCTCTCGAACCCCTTCTCTCTCAACCTTCACGACGCGTCC	323
Db	446	CCCCCCCCCTGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTTCCGAAACGACCTTTTCC	505
Oy	324	GAATCTGAGCTGCGCTTGCGCGCGCGGCGCCACATCCCAAGGCCACAGCAAGAAAGTCG	383
Db	506	CGCGCCATCCCCCCCCCAACACCAACCCCCCCCCCCCCCCCCCCCCCCCCCC	565
Oy	384	ACCTATGTACGAGTCCAGCGGC	407
Db	566	CCCCCATTACTTTTCCACCTGC	589

```

RESULT 14
US-09-804-682-33
; Sequence 33: Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MIMULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
;

```

```
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14, 21, 27, 33, 36, 42, 72, 101, 103, 207, 208, 221, 223,
; LOCATION: 227, 235, 240, 242, 243, 247, 248, 259, 263, 269, 273, 278,
; LOCATION: 296, 321, 322, 324, 330, 332, 335, 336, 340, 371, 385,
; LOCATION: 390, 393, 399, 401, 407, 415, 421, 447, 454, 471, 475
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 479, 494, 501, 508, 511, 513, 525, 539, 540, 579, 596, 605,
; LOCATION: 623, 624, 633, 653, 654, 656, 664, 670, 679, 711, 713,
; LOCATION: 721, 729, 733, 743, 747, 748, 754, 776, 778, 779, 780, 783,
; LOCATION: 798, 802, 808, 810, 813, 814, 820, 822, 824, 825, 838
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 847, 847, 888, 900, 909, 910, 916, 926, 927, 943, 948, 962,
; LOCATION: 973, 1002, 1005, 1028, 1029, 1034, 1057, 1065
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-33
```

```
Query Match 17.9%; Score 93.6; DB 9; Length 1065;
Best Local Similarity 48.4%; Pred. No. 8.5e-14;
Matches 166; Conservative 0; Mismatches 198; Indels 0; Gaps 0;
```

```
QY 27 CTCACACTCCACCTCCAGCGGAGCTGAGCAGAGAGCTCCACCGCGCGGCGCACATG 86
DB 308 CTCACACTCCACCTCCAGCGGAGCTGAGCAGAGAGCTCCACCGCGCGGCGCACATG 367
QY 87 AGTTTCACACCGCTTAACGGCCACCTCTCCAGAGAGCTTCTTCCCTCCAGCGGCG 146
DB 368 CTGNCACCTCTCTTCTTCCAGCGGAGCTGAGCAGAGAGCTCCACCGCGCGGCGCACATG 427
QY 147 TCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 206
DB 428 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 487
QY 207 GCGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 266
DB 488 TTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 547
QY 267 ACCACTAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 326
DB 548 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 607
QY 327 CTGAGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 386
DB 608 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 667
QY 387 TATGTACGAGTGTACGAGCGGCGG 410
DB 668 CCNCTTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 691
```

```
RESULT 15
US-09-804-682-29
; Sequence 29, Application us//09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
```

```
LENGTH: 1064
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5, 6, 16, 21, 24, 25, 33, 39, 72, 110, 209, 214, 231, 232,
; LOCATION: 235, 237, 238, 244, 245, 246, 256, 262, 292, 297, 308, 319,
; LOCATION: 321, 323, 330, 334, 340, 349, 354, 355, 363, 372, 376, 378,
; LOCATION: 397, 405, 432, 437, 454, 455, 457, 458, 459, 468, 470
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 485, 487, 488, 494, 496, 499, 511, 524, 527, 552, 557, 562,
; LOCATION: 583, 600, 611, 613, 623, 624, 652, 654, 674, 681, 688, 691,
; LOCATION: 694, 701, 713, 716, 720, 721, 725, 731, 734, 735, 737, 743,
; LOCATION: 744, 781, 782, 785, 789, 799, 803, 821, 823, 847, 852
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 858, 878, 884, 886, 896, 897, 901, 917, 926, 932, 939, 948,
; LOCATION: 957, 961, 965, 981, 991, 993, 1001, 1002, 1005, 1011, 1018,
; LOCATION: 1043, 1047, 1049, 1051, 1054, 1056
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-29
```

```
Query Match 17.6%; Score 92.2; DB 9; Length 1064;
Best Local Similarity 49.7%; Pred. No. 1.9e-13;
Matches 169; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
```

```
QY 32 CACTCCACTTCACGAGCGAGCTGAGCAGAGAGCTCCACCGCGCGGCGCACATGAGTTG 91
DB 472 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 531
QY 92 TCACACCGTTTACGGCCACCTCTCCAGAGAGCTTCTTCCCTCCAGCGGCGGCTCTG 151
DB 532 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 591
QY 152 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 211
DB 592 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 651
QY 212 CGACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 271
DB 652 NCNCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 711
QY 272 TAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 331
DB 712 TNCNCCNCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 771
QY 332 GCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 371
DB 772 TCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 811
```

```
Search completed: April 14, 2004, 14:27:09
Job time : 305 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 14:05:59 ; Search time 2717 Seconds

(without alignments)
5748.222 Million cell updates/sec

Title: US-09-669-817a-4

Perfect score: 523
Sequence: 1 actcattcagcagatcacca.....gaccgcaagtcgaagtcgga 523

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_hiv:*
19: em_ges_pin:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vr1:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	454.6	86.9	461	9	AUI173524 AUI173524
2	205.4	39.3	527	10	BF421014 BF421014
3	203.8	39.0	601	14	CD463955 ETH1 47 D
4	192.6	36.8	591	14	CA29362 wd11c.pko

5	191.6	36.6	518	14	CA231146	CA231146	SCQSF1303
6	191.2	36.6	843	28	CC345900	CC345900	OGPA181TH
7	189.6	36.3	564	13	BU092717	BU092717	946157A06
8	189.6	36.3	604	14	CA831821	CA831821	111702480
9	172.8	33.0	509	14	CD429988	CD429988	ETH1 16 D
10	162.2	31.0	728	13	BQ762448	BQ762448	EB1001_50
11	160	30.6	483	12	BG356257	BG356257	EMI_23_D0
12	153.6	29.4	658	14	CD213737	CD213737	HS1_52_G0
13	152	29.1	662	14	CA168889	CA168889	SC1FSB101
14	150.4	28.8	657	14	CA217366	CA217366	SC1FSB101
15	150.4	28.8	657	14	CA220794	CA220794	SCBFL405
16	148.8	28.5	491	13	BU099577	BU099577	WHB3309_B
17	147.6	28.2	830	29	CC656592	CC656592	OGPA181TV
18	144	27.5	978	14	CA226699	CA226699	SCVPFL304
19	142.6	27.3	682	13	CA091304	CA091304	SCUTAM200
20	141.8	27.1	686	14	CA212362	CA212362	SCSFAD111
21	141.4	27.0	416	13	BU197746	BU197746	946163H05
22	140.2	26.8	655	13	CA067901	CA067901	SCQSDAD105
23	139.6	26.7	718	14	CA209061	CA209061	SCBZAD1C0
24	135.4	25.9	371	28	BZ590427	BZ590427	3590_1_77
25	133.2	25.5	749	14	CA239979	CA239979	SCSBFL406
26	131.6	25.2	838	9	AA979751	AA979751	MEB1-D3.
27	131.4	25.1	489	14	CA237482	CA237482	SCREFL504
28	131	25.0	789	14	CA224948	CA224948	SCCAGM2C1
29	130.4	24.9	527	28	BH408147	BH408147	1007056D1
30	130.4	24.9	543	14	CF646081	CF646081	3530_1_10
31	130.4	24.9	591	28	BZ584591	BZ584591	3590_1_58
32	130.4	24.9	1607	11	AY103721	AY103721	Zea may
33	130.2	24.9	790	14	CB653015	CB653015	OSJNB03K
34	128.8	24.6	496	12	BM099211	BM099211	EBes01_SQ
35	123.4	23.6	489	12	BJ303004	BJ303004	BJ303004
36	122	23.3	447	13	BU988282	BU988282	HI10701T
37	117.6	22.5	413	12	BI098190	BI098190	IP1_29_CO
38	117.6	22.5	461	12	BI098181	BI098181	IP1_29_BO
39	116.4	22.3	609	13	CA083273	CA083273	SCBPAW201
40	114.8	22.0	530	14	CD568421	CD568421	3529_1_11
41	113.6	21.7	697	14	CD670649	CD670649	OSMR064_5
42	111.2	21.4	844	14	CB907084	CB907084	tr1c081xe
43	111.2	21.3	843	28	AZ200896	AZ200896	SP_1011_A
44	111	21.2	992	29	CNS04R5Q	CNS04R5Q	AL303479
45	110.2	21.1	1292	13	BQ070431	BQ070431	AGENCOURT

ALIGNMENTS

RESULT 1
AUI173524
LOCUS
DEFINITION AUI173524 Rice root Oryza sativa (japonica cultivar-group) cDNA
clone R3620, mRNA sequence.
ACCESSION AUI173524.1 GI:12623311
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 461)
Sasaki, T. and Yamamoto, K.
Rice cDNA from root (2001)
Unpublished (2001)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel.: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@affrc.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT 'RGP'
Location/Qualifiers
1..461

FEATURES
source

Db 297 TCGCCCTCCCTCCCGCAGCCTTCTCTCCACGACTCAGGACCGGCTGGCGGCC 356
 QY 328 TCGAGCTCGCCCTCGCGCGCCCGCCACACCCACAGCCACAGAGAGTGCACCT 387
 Db 357 TGGAGCTCG 413
 QY 388 ATGACACGAGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
 Db 414 CGGCGACGAGCG 473
 QY 448 GCGAGAGGCTCTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 507
 Db 474 GGGCG 533
 QY 508 GCAAGTGAAGTGGGA 523
 Db 534 GCAAGTGGCGTGGGA 549

RESULT 5
 CA231146
 LOCUS 518 bp mRNA linear EST 25-SEP-2003
 DEFINITION SCQSFJ031G02.g Saccharum officinarum FL3 Saccharum officinarum
 cDNA clone SCQSFJ031G02 5', mRNA sequence.
 CA231146
 VERSION CA231146.1 GI:35295392
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Saccharum.
 1 (bases 1 to 518)
 Vector: A.L. da Silva, F.R., Kemper, E.L. and Arruda, P.
 The libraries that made SUCSEST
 Gener. Mol. Biol. 24 (1-4), 1-7 (2001)
 Contact: Arruda, P.
 Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089
 Email: patricia@unicamp.br
 Clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bccc.net.br/br.unesp.br
 Plate: 031 row: G column: 02
 Seq primer: T7 Promoter Primer.
 Location/Qualifiers

FEATURES
 source
 1..518
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCQSFJ031G02"
 /lab_host="DH10B"
 /clone_lib="Saccharum officinarum FL3"
 /note="Organ: Base of developing inflorescence (5cm-long);
 Vector: pSport1. Site 1: SalI. Site 2: NotI. An
 unidirectional cDNA library generated from [base of
 developing inflorescence (5cm-long)]. cDNA was prepared
 from polyA+ mRNA using Superscript Plasmid System Kit
 (Invitrogen). The double-strand cDNAs were fractionated
 in a sepharose CL-2B 40cm-columns and fragments sizing
 between 0.8 and 1.5 kb were directionally cloned into the
 vector. Details of each source of RNA and library
 construction can be obtained at
 http://sucsest.fad.unicamp.br/public"

ORIGIN
 Query Match 36.6%; Score 191.6; DB 14; Length 518;
 Best Local Similarity 68.8%; Pred. No. 9.6e-18;
 Matches 317; Conservative 0; Mismatches 120; Indels 24; Gaps 3;

QY 80 CACCATGAGTCTACACACCGTTAGCGCCACCTCTGAGAGACCCCTTCTCCCTTCC 139
 Db 58 GCGCATGAGCTTCCACACCGCATCTCTCTCGCTCTGAGAGACCCCTTCTCCCTTCC 117
 QY 140 ACCGCG---TCCT 178
 Db 118 ACCGCGCGACCTTCT 177
 QY 179 GCGCTTCTCTCACT 238
 Db 178 CCGCTTCT 237
 QY 239 CCATCTCACCCCTTCT 298
 Db 238 GCGCTTCT 297
 QY 299 CCAACCTTCAACGACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 358
 Db 298 CCGGACCTTCAACGACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 357
 QY 359 CCCCAGG---CCACACGACGAGAGTGAAGTATGATGACGAGTCCACCGCGCGCATGT 415
 Db 358 CCGGAGCG 417
 QY 416 CAAGTGACCAACGAGGAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 475
 Db 418 CAAGTGACCT 477
 QY 476 GCTGACTCCCT 516
 Db 478 GATCAAGTCTCCCGACGACGACGCGCTTCAAGTGGAGGCGGA 518

RESULT 6
 CC345900
 LOCUS 843 bp DNA linear GSS 16-MAY-2003
 DEFINITION OGP181B1TH ZM 0.7_1.5_KB Zea mays genomic clone ZMMBMA0350M17,
 genomic survey sequence.
 CC345900
 VERSION CC345900.1 GI:30815292
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 843)
 White, A., Quackenbush, J., Van Aken, S., Uterback, T.,
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
 Consortium for Maize Genomics
 Unpublished (2002)
 Contact: Cathy White
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: white@tigr.org
 Seq primer: TR
 Classes: sheared ends.
 Location/Qualifiers

FEATURES
 source
 1..843
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBMA0350M17"
 /note="Vector: pBESK-; Site 1: HincII; 0.7-1.5 kb
 methylation filtered genomic DNA library"

ORIGIN
 Query Match 36.6%; Score 191.2; DB 28; Length 843;

Best Local Similarity 67.1%; Pred. No. 1.2e-17;
Matches 306; Conservative 0; Mismatches 138; Indels 12; Gaps 2;

QY 80 CACCATGAGTTCTACACCGGTTACCGCCACCTCTCTGACGAGACCCCTTCTCCCTCC 139
Db 222 CGCCATGGGCTCCACATGCGACATCTCTAGGCTCTCTGAGAGCCCTTCTCCCTCC 281
QY 140 ACCGCGCT 199
Db 282 ACCGCGCGACCT 341
QY 200 ATTCTTGGCCCT 259
Db 342 GTCCCGCT 401
QY 260 CACGCGACCT 310
Db 402 GCCGCGCT 461
QY 311 CGACCGCT 367
Db 462 CACCGCT 521
QY 368 CACGCGAGAGTCT 427
Db 522 CGCCACCGGAGCGACCTATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 581
QY 428 CGAGGACAGCT 487
Db 582 CGTCGAGAGCG 641
QY 488 CTACGAGATGATTCGACCGGAGTGAAGTGGGA 523
Db 642 CGAGGAGAGCGCTTCT 677

RESULT 7
BU092717 564 bp mRNA linear EST 29-AUG-2002.
LOCUS 946157A06.y1 946 - taesael primordialium prepared by Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION BU092717 GI:22542279
VERSION BU092717.1 GI:22542279
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 564)
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946157 row: A column: 06.
Location/Qualifiers
1..564
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="OH43"
/db_xref="taxon:4577"
/tissue_type="taesael"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="X10LR"
/clone_id="946 - taesael primordialium prepared by Schmidt

lab"
/note="Organ: taesael; Vector: HybridAP, Site: 1: EcoRI;
Site: 2: XhoI; George Chuck dissected immature taesael
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridAP. Sample insert size range was 350 bp
to 3 kb with a 1 kb average."

Query Match 36.3%; Score 189.6; DB 13; Length 564;
Best Local Similarity 66.9%; Pred. No. 1.9e-17;
Matches 305; Conservative 0; Mismatches 139; Indels 12; Gaps 2;

QY 80 CACCATGAGTTCTACACCGGTTACCGCCACCTCTCTGACGAGACCCCTTCTCCCTCC 139
Db 17 CGCCATGGGCTCCACATGCGACATCTCTAGGCTCTCTGAGAGCCCTTCTCCCTCC 76
QY 140 ACCGCGCT 199
Db 77 ACCGCGCGACCT 136
QY 200 ATTCTTGGCCCT 259
Db 137 GTCCCGCT 196
QY 260 CACGCGACCT 310
Db 197 GCCGCGCT 256
QY 311 CGACCGCT 367
Db 257 CACCGCT 316
QY 368 CACGCGAGAGTCT 427
Db 317 CGCCACCGGAGCGACCTATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 376
QY 428 CGAGGACAGCT 487
Db 377 CGTCGAGAGAGCG 436
QY 488 CTACGAGATGATTCGACCGGAGTGAAGTGGGA 523
Db 437 CGACGAGAGCGCTTCT 472

RESULT 8
CA831821 604 bp mRNA linear EST 12-DEC-2002
LOCUS 1117024B07.y1 1117 - Unigene V from Maize Genome Project Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION CA831821 GI:26559586
VERSION CA831821.1 GI:26559586
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 604)
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1117024 row: B column: 07.
Location/Qualifiers
1..604

FEATURES
source
1..604
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="OH43"
/db_xref="taxon:4577"
/tissue_type="taesael"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="X10LR"
/clone_id="946 - taesael primordialium prepared by Schmidt

/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="1117 - Unigene V from Maize Genome Project"
/note="This library represents the unique genes found in the fifth round of EST sequencing at Stanford University for the maize genome project. Sequences are present from library 946. Contigs were assembled using ZmBasssembler and 2 representatives from each contig were selected for the Unigene set. All singletons were also selected."

ORIGIN

Query Match 36.3%; Score 189.6; DB 14; Length 604;
Best Local Similarity 66.9%; Pred. No. 1.9e-17;
Matches 305; Conservative 0; Mismatches 139; Indels 12; Gaps 2;

QY 80 CACCATGAGTTCACACCGTTAGCGCCACCTCTCTGACGACGCCCTTCTCCCTCC 139
DB 61 CGCATGAGTTCACACCGTTAGCGCCACCTCTCTGACGACGCCCTTCTCCCTCC 120
QY 140 ACCGCGCT 199
DB 121 ACCGCGCGCACCT 180
QY 200 ATTCTTGGCCCTGACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 259
DB 181 GTCCCGCT 240
QY 260 CACGCGCACCACTAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310
DB 241 GCGCGCGCTCTGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
QY 311 CGACCGGCT 367
DB 301 CACGCGTGTGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
QY 368 CACGCGAGGAGTGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 427
DB 361 CGCCACCGGAGGAGTGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 428 CGAGGAGGAGGAGTGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 487
DB 421 CGTGAAGAGGAGGAGTGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
QY 488 CTACGAGATGATTCGACCGCAAGTGAAGTGA 523
DB 481 CGACGAGAGGAGTGAAGTGAAGTGAAGTGA 516

RESULT 9 CD429988 509 bp mRNA linear EST 03-JUN-2003
LOCUS ETH1.16.D10.G1.A002 Ethylene-treated seedlings Sorghum bicolor cDNA
DEFINITION CD429988
ACCESSION CD429988.1 GI:31345631
KEYWORDS EST.
ORGANISM Sorghum bicolor (sorghum)
SOURCE Sorghum bicolor
KEYWORDS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 509)
Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua, T.N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O., Eastman, A. and Pratt, L.H.
An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid (ACC)-treated seedlings
Unpublished (2003)
JOURNAL OTHER_ESTS: ETH1.16.D10.b1.A002
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: Sugs (CTTCGCTCTAAAGCTCCG).
Location/Qualifiers
1..509
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="ETH1.16.D10.A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. At 8 days of age, medium was supplemented with 0.5 mM 1-aminocyclopropane-1-carboxylic acid (ACC) to induce endogenous ethylene (ETH) production. Roots and shoots were harvested after 27 and 72 hr and material from both time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGG, 3-prime DraIII site is CACCATG). XhoI excises the cDNA insert."

ORIGIN

Query Match 33.0%; Score 172.8; DB 14; Length 509;
Best Local Similarity 68.0%; Pred. No. 4.3e-15;
Matches 291; Conservative 0; Mismatches 122; Indels 15; Gaps 3;

QY 80 CACCATGAGTTCACACCGTTAGCGCCACCTCTCTGACGACGCCCTTCTCCCTCC 139
DB 82 CGCATGAGTTCACACCGTTAGCGCCACCTCTCTGACGACGCCCTTCTCCCTCC 141
QY 140 ACCGCG-----TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 193
DB 142 ACCGCGCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 201
QY 194 CTGCCATTTCTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 247
DB 202 GTCCCGTCTTCCCGACCTGACGACCTCTCTCTCTCTCTCTCTCTCTCTCT 261
QY 248 CCCCTTCTCTCCCGACCGACCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 307
DB 262 CCCCTTCT 321
QY 308 CACGAGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 365
DB 322 CACGAGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 381
QY 366 -CCGACGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 424
DB 382 CCCCGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 441
QY 425 CACGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 484
DB 442 CTCTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 501
QY 485 CCCCTAGC 492
DB 502 CCCGACG 509

RESULT 10
LOCUS B0762448
DEFINITION B0762448 728 bp mRNA linear EST 26-JUL-2002
EBroot1_S0005_J23 R root, 3 week, hydroponic grown, no treatment,
cv Optic, EBroot1 Hordeum vulgare subsp. vulgare cDNA clone
B0762448
ACCESSION B0762448
VERSION B0762448.1 GI:21970920
KEYWORDS EST
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 728)
AUTHORS Hedley, P., Liu H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsey, L., Machray, G., Marshall, D.F.M. and Maugh, R.
JOURNAL Unpublished (2001)
COMMENT Contact: Maugh R, Marshall DP
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scrl.sari.ac.uk.
Location/Qualifiers
1..728
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Optic"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="EBroot1_S0005_J23"
/issue_type="root"
/dev_stage="3 week"
/lab_host="DH103"
/clone_id="root, 3 week, hydroponic grown, no treatment,
cv Optic, EBroot1"
/note="Vector: pSPORT1; Site 1: Sal I; Site 2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old hydroponically grown
unstressured barley plants. Developed as part of the barley
transcriptome resources of BSRC/SEBRAD funded cereal IGF
(Investigating Gene Function) project."

ORIGIN
Query Match 31.0%; Score 162.2; DB 13; Length 728;
Best Local Similarity 63.7%; Pred. No. 1.4e-13;
Matches 300; Conservative 0; Mismatches 158; Indels 13; Gaps 3;
QY 66 CACACGGCGCGGCGACCAAGATTCTCACACCACTTACGCGCACTCTCGACGACCCC 125
DB 40 CCGCGCGCGCGCGCGACAGCTCTCCACCGCGCACTATGCTGCTGACGACGCGC 99
QY 126 TTCTTCCCTCTCCACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
DB 100 TTCTTCCCTCTCCCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 159
QY 186 TCTCTCACTGCGCAT-----TCTTGCGCTTGACTCGCCCTTGCTGTGACCT 236
DB 160 CCACTCTCACTGATCCCATTCGCTTCCCGACGACGACCTCTCTGACCTTACCTC 219
QY 237 TTCCATCTCAACCCCTCTCTCCCGACGCGACCACTAGCTCCCTCTCTGACCCCTTCTC 296
DB 220 CCGCGCGACCCCTTCTTCTCTCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGC 279
QY 297 CTGCAACCTTACCGACGCGCGCTTCTCCAACTGAGTGGCGCTGCGCGCGCGCGCGCA 356
DB 280 TTCTCTCTCTCAGACCTTACCGACGCGCGCTGAGCTTCCGCGCGCGCGCGCGCG 339
QY 357 CACCGCGCGCGCGACGACGAGAGTGAAGTATGTATGACGAGTGCAC---CGCGCGCATG 413

DB 340 GCCCGGACGCGCGCGCGCGACGATACCTACCGCGCGCGCGCGCGCGCGCGCGCAAG 399
QY 414 GTCAAGTGGACACCGACGACACCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 472
DB 400 GTGAAGTGGAGCG 459
QY 473 TGAGCTGACTCCCGCTACGACGATGATTGACCGCGCAAGTGAAGTGGGA 523
DB 460 GAGCTCGCTCCCG 510

RESULT 11
LOCUS BG356257
DEFINITION EML 23 D03.b1_A002 Embryo 1 (EML) sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BG356257
VERSION BG356257.1 GI:13238243
KEYWORDS EST
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 483)
AUTHORS Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
JOURNAL An EST database from Sorghum: developing embryos
Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: emprat@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 475
POLYA=No.
Location/Qualifiers
1..483
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_id="Embryo 1 (EML1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

ORIGIN
Query Match 30.6%; Score 160; DB 12; Length 483;
Best Local Similarity 67.8%; Pred. No. 2.7e-13;
Matches 274; Conservative 0; Mismatches 115; Indels 15; Gaps 3;
QY 80 CACCATGAGTCTACACAGCGTTAGCGCCAGCTCTCTGAGAGACCCCTTCCCTTCC 139
DB 80 GCGATGAGCTTCCACAGCGCACTTCTCGCGCTCTCTGACACACCTTCTTCCCTTCC 139
QY 140 ACCGCG-----TCTCTCTCTCTCTCGCGCTCTCTCTCTCTCTCTCTCTCTCT 193
DB 140 ACCGCGGCACTTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 199
QY 194 CTGCGCATTTCTTGCGCTGAGCTCGCGCT-----CGCTGTACACCTTTCATCTCA 247
DB 200 GTCCCGGTTCCCGCGCGAGCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 259
QY 248 CCGCTTCTCCCGCGCGCGACCACTAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 307

Db 260 CCCCTTCGCTTCCCGCACCCCGTCCCTACATCGTCTCTCCGACCT 319
 QY 308 CACGACCGCGCTCCCACTCGAGCTCGCGTCCGCGCGCCCGACACCCAG-- 365
 Db 320 CACGACCGCGCTCCCGCGCTCGAGCTCGCGTCCGCGCGCGCCCGACACCCAG-- 379
 QY 366 -CCGACGACGAGAGTGCATCTATGTCACCGAGTCCACCGCGCGTGTCAAGTGAC 424
 Db 380 CCGCGCGACCGGAGGACCTACGTACCGACCGCGCGCGCGGAGGAGTCAAGTGAC 439
 QY 425 CACGAGACGAGAGCTCGCGCGCGCGGAGAGGAGTCTCAAGTGAG 468
 Db 440 CTCTGTGAGAGCGCGCGCGCGGAGAGCGTCAAGTGAG 483

RESULT 12

CD213737

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CD213737 658 bp mRNA linear EST 20-MAY-2003
 HSI_52_G09_g1_A012 Heat-shocked seedlings Sorghum bicolor cDNA
 clone HSI_52_G09_A012 5', mRNA sequence.
 CD213737.1 GI:30952107
 EST.
 Sorghum bicolor (sorghum)
 Burkholderia viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 1 (bases 1 to 658)
 Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Buchanan, C.D.,
 Eastman, A. and Pratt, L.H.
 An EST database from Sorghum: heat-shocked seedlings
 unpublished (2003)
 Other ESTs: HSI_52_G09_b1_A012
 Contact: Cordomier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@pratt.uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
 the Human Genome Center, University of Tokyo Institute of Medical
 Science; plant material and RNA prepared at Texas A & M University;
 sequencing done in the Laboratory for Genomics and Bioinformatics,
 University of Georgia. Sequence ends have been trimmed to exclude
 vector and regions below phred quality 16. Three-prime sequences
 are presented as their reverse complement and have been trimmed to
 exclude polyA.
 Seq primer: Sugs (CTTCTGCTCTAAGCTCGC).

FEATURES

source

1..658
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /mol_type="mRNA"
 /cultiVar="TS3620C"
 /db_xref="taxon:4558"
 /clone="HSI_52_G09_A012"
 /lab host="DH10B-TI phage-resistant E. coli"
 /clone.lib="Heat-shocked seedlings"
 /note="Vector: pME18S-FL3, Site 1: XhoI; Site 2: XhoI; The
 library was prepared from polyA+ RNA from 6-day-old
 seedlings grown in hydroponic culture and heat-shocked at
 40-42 C for 4 or 24 hr. After heat shock, roots and leaves
 were harvested and tissues combined for RNA isolation.
 Double-stranded cDNA was cloned unidirectionally into
 different DraIII sites of the pME18S-FL3 vector (5-prime
 DraIII site is CACTGTGTG, 3-prime DraIII site is
 CACCATGTG)."

ORIGIN

Query Match 29.4%; Score 153.6; DB 14; Length 658;
 Best local Similarity 67.4%; Pred. No. 2,3e-12;

Matches 267; Conservative 0; Mismatches 114; Indels 15; Gaps 3;
 QY 80 CACCATGAGTCTTACACCAAGCTTACAGCTCTCTCTGAGAGACCCCTTCCCTCC 139
 Db 263 GCGACGAGCTTCCACCGCACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 322
 QY 140 ACCGCGG-----TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 193
 Db 323 ACCGCGGACACTTCT 382
 QY 194 CTGCGCATTTCTTCT 247
 Db 383 GTCCCGGTTCT 442
 QY 248 CCCCTTCT 307
 Db 443 CCCCTTCT 502
 QY 308 CACGACCGCGCT 365
 Db 503 CACGACCGCGCT 562
 QY 366 -CCGACGACGAGAGTGCATCTATGTCACCGAGTCCACCGCGCGTGTCAAGTGAC 424
 Db 563 CCGCGCGACCGGAGGACCTACGTACCGACCGCGCGCGGAGGAGTCAAGTGAC 622
 QY 425 CACGAGACGAGAGCTCGCGCGCGCGGAGAGGAGTCT 460
 Db 623 CTCTGTGAGAGCGCGCGCGCGGAGAGCGTCAAGTGAG 658

RESULT 13

CA168889

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA168889 662 bp mRNA linear EST 24-SEP-2003
 SCUPSBI10D05.g SBI Saccharum officinarum cDNA clone SCUPSBI10D05
 5', mRNA sequence.
 CA168889
 CA168889.1 GI:35090764
 EST.
 Saccharum officinarum
 Burkholderia viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 1 (bases 1 to 662)
 Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
 The libraries that made SUCSEST
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
 Contact: Arruda P
 Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089
 Email: parnuda@unicamp.br
 Clone distribution: clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bccc.unesp.br
 Plate: 010 row: D column: 05
 Seq primer: T7 Promoter Primer.
 Location/Qualifiers
 1..662
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCUPSBI10D05"
 /lab host="DH10B"
 /clone.lib="SBI"
 /note="Organ: Stalk Bark from adult plants; Vector:
 pSport1; Site 1: SalI; Site 2: NotI; An unidirectional
 cDNA library generated from [Stalk Bark from adult
 plants]. cDNA was prepared from polyA+ mRNA using
 Superscript Plasmid System Kit (Invitrogen). The

double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://succest.lad.ic.unicamp.br/public>

ORIGIN

Query Match 29.1%; Score 152; DB 14; Length 662;
Best Local Similarity 62.4%; Pred. No. 3.8e-12;
Matches 283; Conservative 0; Mismatches 145; Indels 24; Gaps 2;

93 CACGACCGTTAAGCCCACTCTGAGACCCCTTCTCCCTCCACCGCGCTCTCC 152
Db CTCGACCGCGTGAAGAGAGCCCTTCTCCCTCTCCCTCCCTCCCTCCCTCC 105
Qy 153 TCCCTCCCGCCCTCTCTCTCCCGCCGCGCTTCTCCCTCCCTCCCTCTCT 212
Db 106 TCGTCTGCGCCCTCACTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCT 165
Qy 213 GACTCCCGCT 272
Db 166 GACGACCGCT 225
Qy 273 AGCTCCCT 311
Db 226 ACCTTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 285
Qy 312 GACCGCGCT 371
Db 286 GACTTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 345
Qy 372 AGCAGAGTGAAGCTATGTACCGAGTCAACCGCGCATGTGAGTGAAC 431
Db 346 CGTGGCAAGTACACTCTGCTCTCGAGCCGACGAGAGTGAAGTGAAC 405
Qy 432 GACAAGCTGCGCGCGCGAGAGGCTCTCAAGTGGAGGCTGAGTCTCCCT 491
Db 406 GACAGGCT 462
Qy 492 GACGATGATTCGACCGCAAGTGAAGTGAAG 523
Db 463 CACGACGCTTCGACCGCAAGTGAAGTGAAG 494

RESULT 14
CA221736 657 bp mRNA linear EST 25-SEP-2003
LOCUS SCRFPL4008G02.5 FL4 Saccharum officinarum cDNA clone SCRFPL4008G02
DEFINITION 5', mRNA sequence.
CA221736
VERSION CA221736.1 GI:35276347
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 657)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parinda@unicamp.br
Clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
<http://www.bcccenter.fcav.unesp.br>
Plate: 008 row: G column: 02
Seq primer: T7 Promoter Primer.

FEATURES

Location/Qualifiers
1..657
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCRFPL4008G02"
/lab_host="DH10B"
/clone_lib="FL4"
/note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSPori1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://succest.lad.ic.unicamp.br/public>

ORIGIN

Query Match 28.8%; Score 150.4; DB 14; Length 657;
Best Local Similarity 62.4%; Pred. No. 6.4e-12;
Matches 282; Conservative 0; Mismatches 146; Indels 24; Gaps 2;

93 CACGACCGTTAAGCCCACTCTGAGACCCCTTCTCCCTCCACCGCGCTCTCC 152
Db 71 CTCGACCGCGTGAAGAGAGCCCTTCTCCCTCTCCCTCCCTCCCTCCCTCC 130
Qy 153 TCCCTCCCGCCCTCTCTCTCCCGCCGCGCTTCTCCCTCCCTCTCTCTCT 212
Db 121 TCGTCTGCGCCCTCACTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCT 190
Qy 213 GACTCCCGCT 272
Db 191 GACGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 250
Qy 273 AGCTCCCT 311
Db 251 ACCTTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310
Qy 312 GACCGCGCT 371
Db 311 GACTTCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
Qy 372 AGCAGAGTGAAGCTATGTACCGAGTCAACCGCGCATGTGAGTGAAC 431
Db 371 CGTGGCAAGTACACTCTGCTCTCGAGCCGACGAGAGTGAAGTGAAC 430
Qy 432 GACAAGCTGCGCGCGAGAGGCTCTCAAGTGGAGGCTGAGTCTCCCTCT 491
Db 431 GACAGGCT 487
Qy 492 GACGATGATTCGACCGCAAGTGAAGTGAAG 523
Db 488 CACGACGCTTCGACCGCAAGTGAAGTGAAG 519

RESULT 15
CA220794 667 bp mRNA linear EST 25-SEP-2003
LOCUS SCBGFL4050G08.5 FL4 Saccharum officinarum cDNA clone SCBGFL4050G08
DEFINITION 5', mRNA sequence.
CA220794
VERSION CA220794.1 GI:35274458
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 667)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

Thu Apr 15 12:50:26 2004

us-09-669-817a-4.rst

Page 10

TITLE	Genet. Mol. Biol.	24 (1-4), 1-7 (2001)
JOURNAL		
COMMENT		

Job time : 2733 sec

TITLE
The libraries that made SUCEST

JOURNAL
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel.: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parvuda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
<http://www.bcccenter.fcav.unesp.br>
Plate: 050 row: G column: 08
Seq primer: T7 Promoter Primer.
location/Qualifiers
1..667

```

/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCRGF4050G08"
/lab_host="DH10B"
/clone_1fb="F14"
/note="Organ: Developed inflorescence and rachis
(20cm-long); Vector: pSPori1; Site 1: SalI; Site 2: NotI;
An unidirectional cDNA library generated from 10 developed
inflorescence and rachis (20cm-long). cDNA was prepared
from poly(A)+ mRNA using Superscript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://sucest.lad.ic.uncamp.br/public"

```

ORIGIN

Query Match	28.8%;	Score 150.4;	DB 14;	Length 667;
Best Local Similarity	62.4%;	Pred. No. 6.4e-12;		
Matches 282;	Conservative	0;	Mismatches 146;	Indels 24;
				Gaps 2

QY 93 CACACAGGTAAAGCCCAACCTCCTCGACAGACACCCCTTTCACGCTTTCACACGCGCGTCTCC 152
Db 82 CTCACAGCGCGTGCACAGACAGACCCCTTCTTCCATCTCTTCCCTTCCCTTCCCTACACACACAC 141
QY 153 TCTCTCTGCCCCCTTCTCTTCCCAACCCGCGCGTTCCTCCACCTGCGCAATCTTGCGCTTC 212
Db 142 TCGTCTGCGCCCTACATCTCCGCGCTCCGCTCCTCGCCGCCACACACGGCTCTCTCTGAC 201
QY 213 GACTCCCCCTTGGCTGTACCTTTCCATCTCCACCCCTTCTCTCCACGCGCGACACT 272
Db 202 GACACACCCCTTCAACCCCACTCTGCGCGCGCTTCACTCTCCCTCTCCCGATGAC 261
QY 273 AGCTCCCTCTCGACCCCTTCTCTCTCGACACCTTCAC-----C 311
Db 262 ACCTTGACCTGCACCTCCTCCTCGCGGCCCGACAGCGCGCGCGCCCGCTGCTGGGCTTC 321
QY 312 GACCGCGCTCTCCCACTCGAGCTGCGCTCTGCGCGCGCGGCCACACCCCAAGCCCAAC 371
Db 322 GACTTCACCCCTTCTCTCTCGACGCGCTCGCGCACACGCGATCTCGCGCCCGCCCCCG 381
QY 372 AGCAGAAGTGCACCTATGTCAACGAGTCCACCGCGCGGATGTCAAGTGCACACCGAG 431
Db 382 CGTGGCAAGTACACTTACGCTCTCGAGGCGCCACGGCAGAAAGTCTAACTGACCGCCGAG 441
QY 432 GACAAAGCTTGCAGCGCGCGAGAGAGGTGCTCAAGTGGAGAGGCTGACTTCCTCCCTAC 491
Db 442 GACAGGCCC---CGCGCGCGACCGGAACCTCAAGTGGGAAAGCGAGCTCAGACGCCCTTAC 498
QY 492 GACGATGAGTTGCACCGCGCAAGTGGAGTGGGA 523
Db 499 CACGACGAGCTTCAGCGCAAGTGGAAAGTGGGA 530

Search completed: April 14, 2004, 16:08:41